

RICE UNIVERSITY

Essays on Labor Supply Dynamics, Home Production, and Case-based Preferences

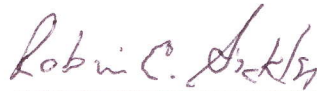
by

Michael Naaman

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

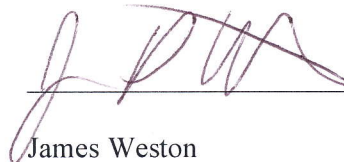
Approved, Thesis Committee:



Robin C. Sickles
Professor Economics, Chair



Richard Boylan
Professor Economics



James Weston
Associate Professor, Jones Graduate
School of Business

HOUSTON, TEXAS
SEPTEMBER 25, 2012

ABSTRACT

Essays on Labor Supply Dynamics, Home Production, and Case-based Preferences

by

Michael Naaman

In this paper we examine models that incorporate CBDT. In the first chapter, we will examine CBDT more thoroughly including a reinterpretation of the standard labor supply problem under a wage tax in a partial equilibrium model where preferences exhibit characteristics of CBDT. In the second chapter, we extend the labor supply decision under a wage tax by incorporating a household production function. Utility maximization by repeated substitution is applied as a novel approach to solving dynamic optimization problems. This approach allows us to find labor supply elasticities that evolve over the life cycle. In the third chapter, CBDT will be explored in more depth focusing on its applicability in representing people's preferences over movie rentals in the Netflix competition. This chapter builds on the theoretical model introduced in chapter 1, among other things, expressing the rating of any customer movie pair using the ratings of similar movies that the customer rated and the ratings of the movie in question by similar customers. We will also explore in detail the econometric model used in the Netflix competition which utilizes machine learning and spatial regression to estimate customer's preferences.

To my parents

Adam Naaman

and

Pamela Naaman

for all their love and support

TABLE OF CONTENTS

1. Making the Case for Case Based Utility Functions.....	1
2. Home Production.....	11
3. The Netflix Competition.....	84
References.....	112

CHAPTER 1: MAKING THE CASE FOR CASE BASED UTILITY FUNCTIONS

In classic economic modeling, economists assume that preferences are an inherent trait of the decision maker. We generally assume that a decision maker knows their preferences over all possible goods and we can represent this preference structure with a quasi-concave, differentiable, utility function.

In certain circumstances, it is natural to develop a model of preferences evolving over time based on past experiences. This is referred to as “case based decision theory” (CBDT) and allows one to model preferences based on recollections of trade-offs, decisions, and the outcomes of those decisions that have played out in the past. Formally, if we let A be the set of acts that are available to the decision maker from some decision problem p and let $c = (a, p, r) \in M$ be the triple consisting of the act, a , chosen in a decision problem, p , and the outcome, r , that resulted from the act, then for any given subset of memories, I , we can express preferences over acts conditional on those memories, which we denote by $\{\succeq_I\}$. If the preference relation satisfies certain axioms we can express the preference relation over acts with a utility function given by

$$U(a) = \sum_{c \in M} s(p, q) u(r) \quad \text{where } c = (a, q, r)$$

So the term $s(p, q)$ is the similarity over the decision problem given that act, a , was chosen. But we may also have a situation where there is similarity between act decision problem pairs. For example, if our set of acts consists of buying or selling a stock and our set of decision problems consists of buying the stock when the price is high or low, then we may have a situation where "buying the stock when the price is low" is more similar to "selling the stock when the price is high" than "selling the stock when the price

is low". Gilboa and Schmeidler (1997) provide axioms that allow a generalization that includes similarity over the pair of decision problems and acts by

$$U(a) = \sum_{c \in M} w((p,a), (q,b)) u(r) \quad \text{where } c=(q,b,r)$$

To see the usefulness of cased based utility models, compare this to the standard approach in a classical partial equilibrium tax model, which will be revisited in chapter 2. Assume we have a representative agent with preferences over consumption, c , and leisure, L , which are both normal goods. The agent has a finite amount of time, which can be used for leisure or labor. Without loss of generality we can normalize the endowment of time to 1, so that h is the percentage of time devoted to labor and $1-h$ is used for leisure. Furthermore, assume that the labor is used in a constant returns to scale technology that produces the consumption good. Suppose there is also a government that levies a labor tax, τ , which will raise revenue, $T = \tau h$. Rogerson (2009) assumes that the government makes a lump sum transfer to the representative agent in the amount of the tax revenue, T . The reason for this assumption is that we are only interested in the substitution effects. By making the lump sum transfer back to the agent, we keep wealth constant. The model can be written as

$$\begin{aligned} \max u(c, L) \text{ st.} \\ c &= (1 - \tau)h + T \\ L &= 1 - h \\ 0 &\leq h \leq 1 \end{aligned}$$

In equilibrium $c = h$, but the agent maximizes utility taking T to be constant. Of course, we must take the leap of faith that the agent takes the lump sum transfer, T , as a constant and doesn't realize that a change in h will produce a change in the lump sum transfer.

However, if we model this tax problem under the guise of a case based utility problem, then there is a much more fluid interpretation. In this case the choice of labor supply, h , will serve as our act. The decision problem in this instance is simply the tax rate that the representative agent faces. Finally the result, r , for any pair (τ, h) is the resulting consumption, c , and leisure, L , which is realized due to the act h .

The maximization problem can be written as a case based decision problem based on memories of different decision problems

$$U(a | p) = \sum_{c \in M} w((h, \tau), (h_i, \tau_i)) u((1 - \tau_i)h_i + T_i, 1 - h_i)$$

If the tax rate changes, the decision maker must decide how to change their labor supply, so the only thing that changes for the decision maker is the tax rate and possibly the labor supply. One can think of this similarity measure as being a kernel that puts extreme weight on the new tax rate, so it acts as an approximate delta function with only results that were close to the new tax rates are kept, so we have

$$\max_h \sum_{\tau_i \approx \tau} s(h, h_i) u((1 - \tau_i)h_i + T_i, 1 - h_i)$$

Since T_i is a result of the act that the government takes to redistribute the tax revenue back to the representative agent, the agent doesn't take into account the fact that $T_i = \tau_i h$ because the tax return is result of an act of another party and not the result of the act taken by the representative agent, so transfers as exogenous. This will prove to be important when savings are included in the model because savings will be considered as transfers from the previous period. As a general rule, it seems reasonable to model experiences that are closer in time as being more relevant which gives

$$U(h, \tau) = \sum_{\tau_i \approx \tau} \beta^i s(h, h_i) u((1 - \tau)h + T_i, 1 - h)$$

If symmetry of the similarity function is assumed, then it is irrelevant whether we maximize over h or over the sequence $\{h_i | 1 \leq i \leq \infty\}$, so an equivalent utility maximization problem is given by

$$\max_{\{h_i | 1 \leq i \leq T\}} \sum_{i=1}^T \beta^i u((1 - \tau)h_i + T_i, 1 - h_i)$$

This gives rise to the standard dynamic optimization problem, but the CBDT forces the restriction

$$1 = \sum_i \beta^i$$

If the main interest of study is the effect of a small change in the tax rate from a rate that has been constant for a very long time, then this approach will be a good approximation to the problem. From this point, the problem can be solved for different tax rates and meaningful comparisons can be made about changes in labor supply. This is a much more natural approach that will be considered empirically in chapter 2.

Consider another example that will motivate our approach in chapter 3. Suppose a decision maker, DM_1 , who has a complete ordering of preferences for N movies.

Assume we ask the person to partition the movies into two groups such that all movies from group G are weakly preferred to group B and we ask them to report a 1 for the movies in group G , and a 0 for the bad movies in group B . Next we take a random draw, z , from the N movies and we ask her to predict the quality of the movie and we give incentives so that she will seek to minimize her prediction error. If we make the draw

with replacement, reporting the proportion of good movies to bad movies will maximize her utility.

Clearly, DM_1 will try to make high quality predictions and the estimate is just the expected utility.

$$EU_1(z) = pU_1(x) + (1-p)U_1(y) = p = \frac{|G|}{N} \quad \text{where } U_1(x)=1 \text{ and } U_1(y)=0$$

In this case, there is no need for the case based utility representation; however, if we make the draws without replacement, then over time DM_1 learns about the distribution of the remaining good movies, so her utility over the precision of the estimates will be maximized at the estimator that minimizes the MSE. At draw $t+1$, she can improve on the naive estimator p with

$$EU_1(z_1^{t+1} | \Omega_1^t) = \frac{|G| - t\bar{u}_1^t}{N-t} = p_1^t$$

where \bar{u}_1^t is the mean utility up to time t and $\Omega_1^t = \{z_1^i, 1 \leq i \leq t\}$

Since we are trying to predict the good movies, we can set the represent this as a weighted average of past cases by using the number of good movies, $|G_1^t|$, and bad

movies, $|B_1^t|$, up to time t , so that $s(p, q_t) = \frac{|G|}{|G_1^t|} - 1$ if the movie is good and

$s(p, q_t) = \frac{|B|}{|B_1^t|} - 1$ if the movie is bad. Of course, the sum of $s(p, q_t)$ up to time t is $N-t$,

we must standardize by the similarity factors, which gives

$$EU_1(z_1^{t+1} | \Omega_1^t) = \frac{|G| - |G_1^t|}{N-t} = p_1^t.$$

This has the nice interpretation as being the conditional expected utility given the past. Also notice that for any time t realization of a good movie increases the bad movie weight and decreases the good movie weight because there is a relatively higher chance of getting a bad movie in the future.

Assume we add another decision maker, DM_2 , with a utility function over movies, U_2 . To keep things simple, assume both parties have access to all of the reported utilities and that both decision makers have the same preference structure. The information set has expanded to include DM_2 , so the conditional expectation will be given by have

$$\begin{aligned} EU_1(z_1^{t+1} | \Omega_1^t \cup \Omega_2^t) &= \frac{|G| - |G_1^t| - |G_2^t| + |G_1^t \cap G_2^t|}{N - t} \\ &= p_1^t + p_2^t - \frac{(|G| - |G_1^t \cap G_2^t|)}{N - t} = \lambda p_1^t + (1 - \lambda) p_2^t \end{aligned}$$

for some $\lambda \in [0, 1]$. if we allow the decision makers to have different preferences, then we can take the correlation, ρ_t , between the two decision makers' ratings, which will be an estimate of the common good movies for both of the decision makers up to time t . For example, if the two decision makers are perfectly correlated, then the second decision maker faces no uncertainty about the quality of the movie. However, we could have perfect correlation up to time t with differing preferences at time $t+1$, so it is not clear how DM_1 would use this imperfect information. If DM_1 just takes a weighted average, then

$$E[E_t[U_1(z_1^{t+1}) | U_2(z_2^{t+1}) = a]] = E[\lambda p_1^t + (1 - \lambda)a] = \lambda p_1 + (1 - \lambda)p_2$$

So our estimate of DM_1 is biased, unless $\lambda = 1$ or $p_1 = p_2$, which violates our assumption. Of course, we could demean the report of DM_2 so

$$E[E_t[U_1(z_1^{t+1}) | U_2(z_2^{t+1}) = a]] = E[\lambda p_1^t + (1 - \lambda)(a - p_2^t)] = \lambda p_1$$

The estimate is still biased unless $\lambda = 1$ and it is unclear how to find λ . However, if we use the best linear predictor based on the estimates of DM_1 about DM_2 , then we will be minimizing the MSE over unbiased estimators, so the utility estimator will be

$$EU_1(z_1^{t+1} | U_2(z_2^{t+1}) = x) = p_1^t + \sigma_1^t \rho_t \frac{(x - p_2^t)}{\sigma_2^t}$$

The second part of this term involves the standardized rating of the second user, the correlation, and the average deviation from the mean of DM_2 . By standardizing the report of DM_2 , we are measuring how much on average DM_2 deviates from her mean. Then we multiply by the correlation, which is our estimate of similarity, and multiply by standard deviation of DM_1 to see how much DM_1 will deviate from her mean given DM_2 .

We can reformulate the problem into the CBDT framework. For any case c we have the act, a , which is the reported rating of movie p and the resulting utility function over acts is given by

$$U(a | p) = \sum_{c \in M} w(a, b) u(r) \text{ where } c = (q, b, r)$$

Unfortunately, we don't know the decision maker's utility, but we do have the reported utility, $\tilde{U}(a | p)$. Presumably, the decision maker is trying to reveal her preferences, so if the decision maker gets utility $u(r, p)$ from act a in decision problem p ,

then we can assume $U(a | p) = f(|u(r) - \tilde{U}(r)|, p)$ where f is some increasing function so that the decision maker maximizes her utility by reporting her utility as truthfully as possible relative to her other reports. If we restrict our attention to cases where the decision maker actually reported a utility, then whenever a is chosen, we have

$$U(a | p) = \sum_{c \in M} s(p, q) \tilde{U}(c)$$

It can be shown that the solution of this maximization problem is to report

$\tilde{U}(p) = E[u(p) | M]$. This means our similarity function should be a probability measure, but the weighting matrix is only unique up to a scale factor. This implies we need to estimate a model of the form

$$(I - \lambda W) \tilde{U} = \varepsilon$$

Of course, that assumes we had the weighting matrix in hand, but we can estimate W assuming we can adequately represent the memories that are being conditioned upon. In order to do this, we use some distance measure, like correlation, and use this distance measure to estimate W with a nonparametric regression. We assume that any distance measure that preserves the similarity ordering will be the same asymptotically, but proving this is beyond the scope of our paper.

Since this person has never faced this decision problem, we would like to gain insight about it. For instance, in our last example, we could set

$$w(a, c) = w(a, (a, t, x)) = \frac{\sigma_1^t \rho_t}{\sigma_2^t}$$

$$I(c) = I(a, t, x) = x - p_2^t$$

Using other peoples reported utility is justified because $I(c)$ represents the information the decision maker has for case c , which could be information about some other decision maker that is relevant. Information about similar decision makers is fundamentally different, so we would like to restrict the cases, c_i , to be positive weights. For example, if a good friend loved some movie, we might take that into account in our decision about what rating to assign that movie. Suppose the movie was a musical and the decision maker hates musicals; but since her friend loved the movie, the decision maker might be inclined to give the movie a higher rating because it was a high quality musical.

$$U(a | I, p) = \sum_{i=1}^n \sum_{c_i \in M} w(a, c) U(b | q) + w(a, c_i) I(c_i)$$

If this new specification can still represent the decision makers' preferences, then it must be the case that our estimating matrix is a linear combination of the true weighting matrix, which means we must scale the estimated weighting matrix.

Our approach, which will be explained in more detail in chapter 3, for scaling the weighting matrix, is to assume the error is normal and perform maximum likelihood on the model

$$(I - \lambda_1 (I \otimes W_1) - \lambda_2 (W_2 \otimes I))U = \varepsilon \text{ where } \varepsilon \sim N(0, I)$$

In order for this model to be identified, W must be invertible. To ensure this is to normalize the rows of W_1 and W_2 one and then restrict $\lambda_1 + \lambda_2 < 1$. This allows us to partition the model into the similar cases for the decision maker and similar decision makers for this decision problem. Since our weights are positive and sum to unity, then

λ_1 and λ_2 can be interpreted as evidence for the expected utility given similar decision problems. Finally, from a statistical perspective, the model is symmetric in movies and users, so *a priori* there is no reason to expect the estimates to be different, but our model shows such a stark contrasts between the models that it gives evidence for the case based utility over similar chosen decision problems.

CHAPTER 2: HOME PRODUCTION

Introduction

Traditionally economists have been very interested in the labor leisure decision choices that consumers make when faced with various tax rates. Prescott (2004) argues that differences in labor tax rates account for the differences in observed labor hours across countries. However, Ragan (2005) finds that countries in continental Europe with low levels of market work also tend to have high levels of time devoted to home production as compared with the US. This result indicates that it might be helpful to examine the problem by taking into account labor utilized in home production as well as market labor in the labor supply problem.

The following table shows the hours worked for various OECD economies relative to US labor hours. This table shows dramatic differences in labor across countries with the economies of Europe working far less than the US. One possible explanation of these differences might be differences in labor tax rates across countries. McDaniel (2006), using data on 15 OECD countries from 1950-2003, finds that the effective average labor tax rate in the countries with the highest market labor supply is around 30% and 50% in the countries with a relatively low market labor supply. We will use these results as our bench mark for comparing the change in market labor supply for a change in tax rates from 30% to 50%.

Country	Average effective labor	hours worked relative to the
---------	-------------------------	------------------------------

	tax rate	US in 2006
Japan	32.6	1.02
New Zealand	40.6	1.00
Canada	40.3	.98
Ireland	44.8	.98
Australia	29.6	.96
Portugal	44.2	.96
Denmark	62.1	.93
Switzerland	40.6	.93
Sweden	63.2	.91
Finland	62.1	.90
Greece	49.9	.90
UK	36.5	.90
Spain	46.0	.88
Austria	47.3	.81
Norway	58.0	.81
Netherlands	59.6	.77
Belgium	57.0	.73
France	57.2	.73
Germany	50.7	.73
Italy	55.4	.70

As we can see from the table, the European countries, with the exception of Ireland, do not work as much as the other OECD countries. However, there is heterogeneity within Europe as well because we can see that many of the Northern European countries, such as Denmark, Sweden, and Finland, have higher tax rates and higher labor hours relative to their Western European counterparts, such as France, Germany, and Italy. It appears that there may be more to these differences than simply tax rates. For example, if we run a simple regression with the data in the table above, gives the following results

$$\ln h_i = 3.247 - .305 \ln \tau_i + \varepsilon_i$$

(.0719) (.0901)

where h_i is the average number of hours worked per week in country i and τ_i is an estimate of the average tax rate in country i . Prescott (2004) argues that tax rates are the main determinant of the differences in labor supply for G7 countries. The simple regression confirms the importance of the tax rate in the role of labor supply, but the poor fit of the model ($R^2 = .35$) indicates that the model lacks explanatory power.

As an alternative to the Prescott model, Rogerson (2009) assumes a representative agent has preferences over consumption, c , and leisure, L , which are both normal goods. The agent has a finite amount of time, which can be used for leisure or labor. Without loss of generality we can normalize the endowment of time to 1, so that h is the percentage of time devoted to labor and $1-h$ is used for leisure. Furthermore, assume that labor is used in a constant returns to scale technology that produces the consumption good. Suppose there is also a government that levies a labor tax, τ , which will raise revenue, $T = \tau h$. Rogerson assumes that the government makes a lump sum transfer to

the representative agent in the amount of the tax revenue, T , in order to examine the substitution effects. By making the lump sum transfer back to the agent, we keep wealth constant. The model can be written as

$$\begin{aligned} \max_{l \geq h \geq 0} u(c, 1-h) \\ \text{st. } c = (1-\tau)h + T \end{aligned}$$

In equilibrium $c = h$, but the agent maximizes utility taking T to be constant. As an example, consider the following maximization problem.

$$\max_{l \geq h \geq 0} \alpha \ln((1-\tau)h + T) + \frac{1-\alpha}{1-\gamma} [(1-h)^{1-\gamma} - 1]$$

The first order conditions are given by

$$\frac{\alpha(1-\tau)}{(1-\tau)h + T} - \frac{(1-\alpha)}{(1-h)^\gamma} = 0$$

Substituting the equilibrium condition $T = \tau h$ into the FOC allows us to establish the following relationship in equilibrium

$$\frac{\alpha(1-\tau)}{(1-\alpha)} = \frac{h}{(1-h)^\gamma}$$

In general there will be no closed solution, but it can be shown that an increase in the tax rate results in a decrease in h and an increase in leisure, which is just $1-h$. The magnitude of these changes will be determined in large part by the agent's taste for leisure, which is controlled by γ . If we restrict our attention to small integer values of γ , then a closed form solution can be recovered for labor supply, which can be seen in the table below

labor supply in equilibrium for a given γ	γ
--	----------

$h^* = \frac{\alpha(1-\tau)}{(1-\alpha)}$	0
$h^* = \frac{\alpha(1-\tau)}{1-\alpha\tau}$	1
$h^* = 1 + \frac{(1-\alpha)}{2\alpha(1-\tau)} \left(1 - \sqrt{1 + \frac{4\alpha(1-\tau)}{1-\alpha}} \right)$	2

As we can see the solution to the maximization problem becomes progressively more complicated as γ increases, but it is interesting to note that without any tax the labor supply is decreasing as γ get larger.

However, it is possible to examine the comparative statics of such a model, first calibrate the model to match the US economy. Following McDaniel (2006) we set the tax rate to .3, furthermore, we assume that in equilibrium $h = \frac{1}{3}$ for the US economy. As a reference, if we assume people average 8 hours of sleep a night, then there will be 119 waking hours a week devoted to labor or leisure and a 40 hour work week is roughly 1/3 of that amount. The specification of these two values allows us to determine the corresponding value of α in order to calibrate the model to have features similar to the US economy.

Rogerson (2009) uses a calibrated version of α to calculate the time in the labor market associated with a tax rate of .5 relative to that in the equilibrium of the US calibrated model with a tax rate of .3. The results are given in the table below.

The Ratio of Labor at the New Equilibrium	γ
---	----------

with a 50% Tax Rate Relative to a 30%	
Tax rate in the Old Equilibrium	
.76	.5
.79	1
.84	2
.9	5
.94	10
.97	20

As we can see from the table, the labor supply becomes more inelastic as γ grows larger.

As an alternative to this approach, we can attempt to calculate the labor supply elasticity directly.

We first redefine the first order conditions as an implicit function of labor supply and tax rates

$$F(\tau, h) = \frac{\alpha(1-\tau)}{(1-\tau)h + T} - \frac{(1-\alpha)}{(1-h)^\gamma} = \frac{\alpha(1-\tau)}{h} - \frac{(1-\alpha)}{(1-h)^\gamma} = 0$$

By the concavity of u , $F_2 \geq 0$ which means we can apply the implicit function theorem

$$\frac{\partial h}{\partial \tau} = \frac{-F_1}{F_2} = \frac{\alpha/h}{(-1)\left[\frac{\alpha(1-\tau)}{h^2} + \frac{\gamma(1-\alpha)}{(1-h)^{\gamma+1}}\right]} = -\frac{1}{1-\tau} \left[\frac{1}{h} + \frac{\gamma}{1-h}\right]^{-1}$$

where the second equality uses the fact that the FOC must satisfy $\frac{\alpha(1-\tau)}{1-\alpha} = \frac{h}{(1-h)^\gamma}$.

The Marshallian labor supply elasticity, e_{market}^M , is given by

$$e_{market}^M(\tau) = \frac{\tau}{h} \frac{\partial h}{\partial \tau} = - \left(\frac{\tau}{1-\tau} \right) \left(\frac{1-h}{1-h+\gamma h} \right)$$

The expression indicates e_{market}^M depends entirely on γ . To see this, notice that

$$\lim_{\gamma \rightarrow 0} e_{market}^M = \frac{-\tau}{1-\tau} \text{ and } \lim_{\gamma \rightarrow \infty} e_{market}^M = 0 \text{ which is consistent with the result in Rogerson}$$

(2009), in that there is very little change in market work as γ increases. This is intuitive as γ controls the importance of leisure to the representative agent. Another advantage of this approach is that we can derive a reasonable estimate for γ .

Based on the simply estimated relationship above we can solve for γ for any country given their tax rate and labor supply. Since

$$\ln h_i = 3.247 - .305 \ln \tau_i + \varepsilon_i$$

(.0719) (.0901)

then since $\frac{\partial \ln h_i}{\partial \ln \tau_i} = -.305 = \frac{\tau_i}{h_i} \frac{\partial h_i}{\partial \tau_i} = e_M$ we can derive a corresponding γ for any

country given. For example, if we assume $h = \frac{1}{3}$ and $\tau = .3$ for the US then assuming our model is correctly specified and countries are homogenous, then $\gamma \approx .75$; however, if we repeat this for all of the countries we find a range from .75 to 11 for implied values of γ . Rogerson (2009) points out that the differences in labor supply between the US and Western Europe could be accounted for by this model if $\gamma \leq 1$, but the implied γ for the western European countries is far greater than 1 as the table below indicates.

Country	Implied γ
Japan	1.07

New Zealand	2.38
Canada	2.4
Ireland	3.3
Australia	0.74
Portugal	3.27
Denmark	9.47
Switzerland	2.65
Sweden	10.35
Finland	9.92
Greece	5.12
UK	1.97
Spain	4.17
Austria	5.07
Norway	9.27
Netherlands	10.80
Belgium	10.11
France	10.21
Germany	7.14
Italy	9.8

We would expect that these implied values of γ should be clustered around the same value of γ , but the vast variation indicates that there is probably some variation amongst different countries.

In the present context leisure is simply all activities that occur outside the market place. This is problematic because any data that we have is inherently based on market information. If we have information on prices and quantities demanded, we can recover the agent's preferences; however, in the case of leisure, the relative price of leisure compared to other goods is unobservable. If we put more structure on the concept of leisure, then the problem becomes more tractable.

Suppose people produce at home a subset of their consumption goods more efficiently than the market place (Becker, 1965). For instance, it is far cheaper to cook a high quality steak at home than purchasing it at a restaurant. However, the keyword is "cheaper"; the consumer knows when it is beneficial to cook the steak as opposed to eating out. Unfortunately, it is very difficult for the econometrician to measure this notion of "cheaper" and thus the econometrician may not be able to measure leisure in its most general context. However, one may still view home production choices as suitable proxy for a substantial portion of the waking day spent away from traditional market-based labor activities.

Several studies offer insight into the differences in labor supply decisions between the US and European countries. Freeman and Schettkat (2005) report that time spent in home production in European countries is about 20% greater than in the US. Burda (2008) reaches a similar conclusion that Europeans spend 15% to 20% more time in home production in addition to differences in leisure of about 15%. Intuitively we would

expect that countries with higher labor tax rates have a smaller market labor supply for those goods which can be more easily produced outside the market, i.e. cooking at home instead of eating out at a restaurant. Studies by Davis and Henrekson (2004) and Freeman and Schettkat (2005) find evidence for this intuition. It seems reasonable to assume that the smaller market labor supply of Europe could be due to increased leisure time in Europe and increased home production, and thus these two factors should be considered separately in any time allocation model.

Adding home production to a model can account for the disparity in labor supply between the US and Europe. Following Becker (1965), we assume there is a home production function, f , that takes as inputs goods, g , and time, h_2 , with an output that is the consumption good so that

$$\begin{aligned} c &= f(g, h_2) \\ g &= (1 - \tau)h_1 + T \\ h_1 + h_2 + L &\leq 1 \end{aligned}$$

where L is leisure time and h_1 is the proportion of time devoted to market labor. Here g plays a similar role as before as the after tax labor revenue plus the lump sum tax that the consumer can use in the home production process along with the home production labor supply, h_2 , to produce the final consumption good.

For example, a meal at home requires some labor income to buy ingredients and time in order to cook a meal at home. As in Rogerson (2009), we assume that the home production process will be given by

$$c = \left(\theta g^\varepsilon + (1 - \theta) h_2^\varepsilon \right)^{1/\varepsilon}$$

where $0 \leq \beta \leq 1$ and $\varepsilon \leq 1$ so that the home production function will be a CES function exhibiting constant returns to scale. The new utility maximization program is thus given by

$$\begin{aligned} \max_{l \geq h_1 + h_2 \geq 0} \quad & \alpha \ln(c) + \frac{1-\alpha}{1-\gamma} \left[(1-h_1-h_2)^{1-\gamma} - 1 \right] \\ \text{st. } \quad & c = \left(\theta[(1-\tau)h_1 + T]^\varepsilon + (1-\theta)h_2^\varepsilon \right)^{1/\varepsilon} \\ & h_1 \geq 0, h_2 \geq 0 \end{aligned}$$

After substituting in the equilibrium condition $T = \tau h_1$, The FOC for this model will be given by the following system of equations

$$\begin{aligned} 0 &= \frac{\alpha \theta (1-\tau) h_1^\varepsilon}{\theta h_1^\varepsilon + (1-\theta) h_2^\varepsilon} - \frac{(1-\alpha) h_1}{(1-h_1-h_2)^\gamma} - \lambda h_1 \\ 0 &= \frac{\alpha (1-\theta) h_2^\varepsilon}{\theta h_1^\varepsilon + (1-\theta) h_2^\varepsilon} - \frac{(1-\alpha) h_2}{(1-h_1-h_2)^\gamma} - \lambda h_2 \end{aligned}$$

where λ is the Kuhn-Tucker multiplier for the constraint on the two types of labor. It can be shown that any solution of this system must be in interior solution with $\lambda = 0$.

Rogerson (2009) utilizes a similar calibration process as before. He assumes a tax rate of .3 to represent the US economy with market labor given by $h_1 = \frac{1}{3}$ and home

production labor given by $h_2 = \frac{1}{4}$. The estimates for home and market labor are

supported by the work of Francis and Ramsey (2007). Once we have these values set for the US economy in equilibrium, we can back out values for α and θ for any given values of γ and ε . This allows us to examine the change in market labor, home labor,

and leisure when the tax rate is changed from .3 to .5. The tables below give the results from Rogerson (2009)

market labor (h_1)	$\gamma = .5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$
$\varepsilon = 0$.76	.79	.81	.83	.84	.85
$\varepsilon = .4$.69	.71	.73	.74	.75	.75
$\varepsilon = .5$.66	.67	.68	.70	.70	.70
$\varepsilon = .6$.60	.61	.62	.63	.63	.64

We can see from the table above that as γ becomes larger leisure becomes less important and market labor increases for a fixed ε due to the substitution of leisure to market hours. Similarly as ε becomes larger for a fixed γ we see that market labor decreases as there is a substitution from market labor to home labor and that decrease is by much more than the 20% raise in taxes.

home labor (h_2)	$\gamma = .5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$
$\varepsilon = 0$	1.07	1.10	1.14	1.17	1.18	1.19
$\varepsilon = .4$	1.21	1.25	1.27	1.30	1.31	1.32
$\varepsilon = .5$	1.28	1.32	1.34	1.37	1.38	1.38
$\varepsilon = .6$	1.40	1.40	1.44	1.46	1.47	1.47

As one would expect, as γ becomes larger in the table above, leisure becomes less important and home labor increases for a fixed ε . There is a substitution of leisure to home labor and market labor. Similarly as ε becomes larger for a fixed γ we see that now home labor increases as there is a substitution from market labor to home labor and for $\varepsilon > 0$ we see a corresponding increase in home labor supply that is also greater than the 20% increase in taxes.

leisure ($1 - h_1 - h_2$)	$\gamma = .5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$
$\varepsilon = 0$	1.14	1.10	1.07	1.03	1.02	1.01
$\varepsilon = .4$	1.11	1.08	1.05	1.02	1.01	1.01
$\varepsilon = .5$	1.10	1.07	1.04	1.02	1.01	1.01
$\varepsilon = .6$	1.07	1.05	1.03	1.02	1.01	1.01

From the final table in this series, we can see that leisure is inelastic as all changes in leisure are less than the 20% raise in taxes, furthermore, for large γ the change in taxes barely affects leisure independent of ε which means the decrease in market labor goes almost entirely to home labor. This intuitively makes sense because the consumer is using home production to evade the tax levied by the government. This is a simple, but relatively effective approach to the comparative statics problem, however, we can say little more about the problem.

First reconsider the first order conditions from the maximization problem

$$0 = \frac{\alpha\theta(1-\tau)h_1^\varepsilon}{\theta h_1^\varepsilon + (1-\theta)h_2^\varepsilon} - \frac{(1-\alpha)h_1}{(1-h_1-h_2)^\gamma} - \lambda h_1$$

$$0 = \frac{\alpha(1-\theta)h_2^\varepsilon}{\theta h_1^\varepsilon + (1-\theta)h_2^\varepsilon} - \frac{(1-\alpha)h_2}{(1-h_1-h_2)^\gamma} - \lambda h_2$$

It should be noted that when $\varepsilon = 1$, we will have a corner solution, which is given by

$$h_1 = H\left(\frac{\theta(1-\tau)}{1-\theta} - 1\right)$$

where H is the Heaviside step function, so assume $\varepsilon < 1$ for the rest of the analysis.

Since there is a common term for both equations at an interior solution, we can substitute the first equation into the second to get the following relationship

$$1 = \frac{\alpha\theta(1-\tau)h_1^{\varepsilon-1}}{\alpha(1-\theta)h_2^{\varepsilon-1}}$$

This means h_1 and h_2 must satisfy $\frac{h_1}{h_2} = \left[\frac{\theta(1-\tau)}{1-\theta}\right]^{\frac{1}{1-\varepsilon}} \equiv A$ which implies

$h_1 + h_2 = (1+A)h_2 \leq 1$ giving a strict upper bound of $\frac{A}{1+A}$ and $\frac{1}{1+A}$ for h_1 and h_2 ,

respectively. Define $D = 1 + \frac{\theta A^\varepsilon}{1-\theta}$ and solve for h_1 which gives

$$h_1 = \left(\frac{\alpha}{1-\alpha}\right) \frac{L^\gamma A}{1 + \theta A^\varepsilon / (1-\theta)} = \left(\frac{\alpha}{1-\alpha}\right) \frac{L^\gamma A}{D}$$

Market labor is a function of only leisure and the model parameters. We can also find the optimal leisure choice as a function of market labor, which is given by

$$L = 1 - h_1 - h_2 = 1 - (1+A)h_2 = 1 - (1+A^{-1})h_1$$

Unfortunately, there is no closed form solution for the choices of labor and leisure, but the implicit function theorem always allows us to find the derivative of the

demand for labor and leisure with respect to the tax rate. The FOC are satisfied if and only if

$$\begin{aligned} F(h_1, L) &= L - 1 + (1 + A^{-1})h_1 = 0 \\ G(h_1, L) &= h_1 - \left(\frac{\alpha}{1 - \alpha} \right) \frac{L^\gamma A}{D} = 0 \end{aligned}$$

Since these two functions are constant for all tax rates, then their derivatives must be zero, which means we must solve the following system of equations

$$\begin{aligned} \frac{\partial F}{\partial \tau} &= \frac{\partial L}{\partial \tau} + (1 + A^{-1}) \frac{\partial h_1}{\partial \tau} + \frac{\partial A^{-1}}{\partial \tau} h_1 = 0 \\ \frac{\partial G}{\partial \tau} &= \left(\frac{\alpha}{1 - \alpha} \right) \left(\frac{\partial (A/D)}{\partial \tau} L^\gamma + \frac{\gamma A L^{\gamma-1}}{D} \frac{\partial L}{\partial \tau} \right) - \frac{\partial h_1}{\partial \tau} = 0 \end{aligned}$$

After a good deal of algebra, it can be shown that the market labor supply elasticity is given by

$$e_{market}^{MH}(\tau) = \frac{\tau}{h_1} \frac{\partial h_1}{\partial \tau} = \frac{-\tau}{1 - \tau} \left[1 + x - \left(\frac{\gamma(1 - L) \left(1 + x - \frac{1}{(1 - \varepsilon)(1 + A)} \right)}{L + \gamma(1 - L)} \right) \right]$$

where $x = \frac{\varepsilon}{(1 - \varepsilon)D}$ and e_{market}^{MH} is the Marshallian elasticity with home production. Since

x doesn't depend on γ , we can see that the home production parameter, ε , is a first order effect while the leisure parameter, γ , is a second order effect. To see this fact more clearly, notice that for large γ

$$e_{home}^{MH}(\tau) \approx \frac{-\tau}{1 - \tau} \left[\frac{1}{(1 - \varepsilon)(1 + A)} \right]$$

The effect of γ is essentially washed out if it is large and most of the elasticity is determined by the ε parameter.

If we turn our attention to home labor, we can take advantage of the relationship between market labor and home labor writing

$$\frac{\partial h_2}{\partial \tau} = \frac{\partial A}{\partial \tau} h_2 + A \frac{\partial h_1}{\partial \tau}$$

This allows us to express home labor supply elasticity, e_{home}^{MH} , in terms of market labor supply elasticity which gives

$$e_{\text{home}}^{MH}(\tau) = e_{\text{market}}^{MH}(\tau) + \frac{\tau}{(1-\tau)(1-\varepsilon)}$$

As noted previously $\varepsilon < 1$ at an interior solution, so market labor is always more elastic than home labor and the difference between the two elasticities depends only on the tax rate and the home production parameter, ε . This makes sense because a consumer will respond to a tax increase by substituting market labor for home labor and leisure.

The home production model will have a labor supply elasticity at least as large as the model without home production. In general if we are modeling the price elasticity of apples and we add a substitute for apples to the model, then the new price elasticity of apples should be at least as large in the new model that has more substitutes for apples. More formally take the total derivative of the time endowment equation for the home production model

$$0 = \frac{\partial L}{\partial \tau} + \frac{\partial h_1}{\partial \tau} + \frac{\partial h_2}{\partial \tau}$$

In the home production model $\frac{\partial h_2}{\partial \tau} > 0$ as long as $\frac{1}{1-\varepsilon} > \frac{\partial \ln h_1}{\partial \ln w}$ which simply means that that elasticity of substitution between the two types of consumption is greater than the wage elasticity. This is important because it means that

$$-\frac{\partial L}{\partial \tau} < \frac{\partial h_1}{\partial \tau}$$

All other things being equal, home production is a parameter in the standard model which means that $\frac{\partial h_2}{\partial \tau} = 0$, so in the standard model without home production

$$-\frac{\partial L}{\partial \tau} = \frac{\partial h}{\partial \tau} < \frac{\partial h_1}{\partial \tau}$$

The intuition behind this result is that a tax increase in the home production model implies labor must go down to allow home production and leisure to increase. While in the standard model, any decrease in labor goes directly to leisure. Implicitly we have assumed that income effect is the same in both models, which holds for the Hicks elasticity, but not the Marshallian elasticity. So our model predicts that the Hicks elasticity will be larger in the home production model, but income effects can muddle this relation between the two models. This simple model demonstrates the importance of non-market labor. Taxes create a market distortion in the labor market, but they also affect home production and leisure, which need to be taken into account.

Leisure in its most general form cannot be estimated from market data. However, in this simple model, home production will dominate "regular leisure". It is not reasonable to assume that we can measure a person's enjoyment of a nice glass of wine spent with her family. But we can measure the tradeoff between eating at home as

opposed to eating at a restaurant and that tradeoff can be measured by the time devoted to restaurant dining as opposed to home dining.

In order to identify the role of taxes and home production on labor supply, we need to move to a dynamic setting to see the how labor supply and home production responds to changes in effective tax rates. Moving to the dynamic setting, assume that at time t market consumption is given by c_t , market labor supply is given by h_{mt} , market labor supply is given by L_t , and hourly wage given by w_t . Ghez and Becker (1975) present a dynamic labor supply problem where the individual solves

$$\begin{aligned} \max_{(c_t, h_{mt}, L_t) | 1 \leq t \leq T} & \sum_{t=1}^T \beta^t U(c_t, L_t) \\ \sum_{t=1}^T (1+r)^{-t} (c_t - w_t h_{mt}) & \leq I_0 \\ L_t + h_{mt} & \leq H_t \end{aligned}$$

where I_0 is the initial exogenous income, β is the discount factor, r is the interest rate, and H_t is the time endowment. In the static case, we rescaled the time endowment to unity without loss of generality because there is only 24 hours in a day, but now we insist on explicitly modeling the time endowment because we are going to control for basic functions such as sleep and personal care which can vary across individuals and across time. So now H_t will be the time endowment that can be allocated to market labor and leisure.

In the classic work by Ghez and Becker (1975), a synthetic cohort is created and the resulting labor supply equation is given by

$$\ln h_{ct} - \ln h_{ct-1} = \beta + \delta (\ln w_{ct} - \ln w_{ct-1}) + \varepsilon_{it}$$

where $\ln h_{ct}$ is the average of the log of labor supply for all people in the 3-year age cohort. The result of this regression gives

$$\ln h_{ct} - \ln h_{ct-1} = -.002 + .282(\ln w_{ct} - \ln w_{ct-1}) + \varepsilon_{it}$$

Using our data, I run a similar regression, but the table below shows the estimate of δ for all data up to that year, so we can see how the estimate changes as we add more data

year	δ
1976	-0.12875
1985	-0.12923
2003	0.013917
2004	0.184585
2005	0.178054
2006	0.189466
2007	0.212118
2008	0.254941
2009	0.241251
2010	0.258599

The first thing to notice is that it takes about four years of data before the estimate begins to stabilize. In the dynamic setting MaCurdy estimates a labor equation in a 2SLS setting where the wage is treated as endogenous

$$\ln h_{it} - \ln h_{it-1} = \beta + \delta(\ln w_{it} - \ln w_{it-1}) + \varepsilon_{it}$$

In this model the change in wages is considered endogenous, so family background variables, age, education, interaction between age and education, and dummy variables for time are used as instrumental variables. This result of the 2SLS estimation is

$$\ln h_{it} - \ln h_{it-1} = -.009 + .23(\ln w_{it} - \ln w_{it-1}) + \varepsilon_{it}$$

Since MaCurdy had a true panel, we must create a synthetic cohort, but the results are similar

year	δ
1976	-1.58999
1985	-0.16145
2003	0.108653
2004	0.108003
2005	0.082572
2006	0.056722
2007	0.155133
2008	0.184813
2009	0.155247

It is usually assumed in the macroeconomics literature that the utility function is separable in consumption and leisure, which is to say the utility function admits a representation of the form $U = u(c_{mt}) - v(h_{mt})$, further discussion of this representation can be found in Prescott (2006).

For example, Rupert, Rogerson and Wright (2000) set $v(h_{mt}) = \phi h_{mt}^\gamma$ which results in the following dynamic optimization problem

$$\max_{(c_t, h_{mt}, L_t | 1 \leq t \leq T)} \sum_{t=1}^T \beta^t \left[u(c_t) - \phi h_{mt}^\gamma - \lambda \beta^{-t} \left(\frac{c_t - h_{mt}}{(1+r)^t} - I_0 \right) \right]$$

$$L_t + h_{mt} \leq H_t$$

One of the benefits of this functional form is that consumption will not appear in the first order conditions for labor supply which are given by

$$-\beta^t \phi \gamma h_{mt}^{\gamma-1} + \frac{\lambda}{(1+r)^t} = 0$$

Rearranging and taking the logarithm of both sides, we have a feasible estimating equation given by

$$(\gamma - 1) \ln h_{mit} = \ln \left(\frac{\lambda}{\phi \gamma} \right) - t \ln(\beta(1+r)) + \ln(w_{it}) + \varepsilon_{it}$$

Rupert, Rogerson and Wright (2000) only had cross sectional data, so they follow Ghez and Becker (1975) by creating a synthetic cohort that creates a pseudo panel data set. In order to create the synthetic cohort, the relevant variables are averaged for all people of the same age, so our regression equation becomes

$$(\gamma - 1) \ln h_{mat} = \ln \left(\frac{\lambda_a}{\phi \gamma} \right) - a \ln(\beta(1+r)) + \ln(w_{at}) + \varepsilon_{at}$$

where $\ln h_{mat}$ is the mean of the logarithm of hours of market work for all individuals of age a in period t . Also note that a has replaced t because age represents the number of periods that the individual has been solving the dynamic optimization problem, but it is a bit unreasonable to assume that individuals have been optimizing since birth, so age in this context begins when the first cohort is created which is 22 in this case.

It is further assumed that the data generating process follows a balanced growth path. The balanced growth path assumption means that

$$c_{t+1} = (1+g)c_t$$

$$\lambda_{a+1} = \frac{\lambda_a}{1+g} = \dots = \frac{\lambda_0}{(1+g)^a}$$

where g is the growth rate of consumption. This allows the regression equation to be written as

$$\ln h_{mat} = \frac{1}{\gamma-1} \ln \left(\frac{\lambda_0}{\phi^\gamma} \right) - \frac{a}{\gamma-1} \ln [\beta(1+r)(1+g)] + \frac{1}{\gamma-1} \ln(w_{at}) + \frac{1}{\gamma-1} \varepsilon_{at}$$

Unfortunately, this model is not identified when $\gamma = 1$ which is a relevant value, but they do find that the Frisch labor supply elasticity is 0.146.

In order to capture a more realistic model of labor decisions, the model is then extended to include a home production in a similar fashion as our paper, so that the individual will solve

$$\begin{aligned} \max_{(c_t, h_{mt}, h_{ht}, I_0) | 1 \leq t \leq T} \sum_{t=1}^T \beta^t [u(c_t) - \phi(h_{mt} + h_{ht})^\gamma] - \lambda \left(\frac{c_t - h_{mt}}{(1+r)^t} - I_0 \right) \\ h_{ht} + h_{mt} \leq H_t \end{aligned}$$

where h_{ht} is the number of hours devoted to household activities. This specification allows for the disutility of home production, but it does not address the role of home production in the production process of consumption. This model captures the disutility of cooking dinner at home, but it ignores the fact that cooking is part of the production process of consuming dinner at home.

Nevertheless, the estimating equation for the home production model is similar to the standard model. It is given by

$$\ln(h_{amt} + h_{aht}) = \frac{1}{\gamma-1} \ln \left(\frac{\lambda_0}{\phi^\gamma} \right) - \frac{a}{\gamma-1} \ln [\beta(1+r)(1+g)] + \frac{1}{\gamma-1} \ln(w_{at}) + \frac{1}{\gamma-1} \varepsilon_{at}$$

In general, λ_0 will be a function of the parameters including the wage, so

$$\frac{\partial \ln(h_{mt})}{\partial \ln w_t} = \left(\frac{\ln(\lambda_0)}{1-\gamma} \frac{\partial \lambda_0}{\partial \ln w_t} + \frac{1}{1-\gamma} \right)$$

The first term in the expression above represents the change in labor supply due to the income effect and the second term is the change due to the substitution effect. Rupert, Rogerson and Wright (2000) assume that λ_0 is exogenous and constant across all individuals, which is restrictive, but reasonable for their purposes because they were mainly interested in the substitution effect. Once home production is added to the model, they find that the Frisch elasticity is 0.297.

Another model that was estimated is similar to our model in that the individual solves the slightly different dynamic optimization problem, which is given by

$$\begin{aligned} \max_{(c_t, h_{mt}, h_{ht}, I_0 | 1 \leq t \leq T)} & \sum_{t=1}^T \beta^t \left[u(c_t) + \phi(H_t - h_{mt} - h_{ht})^\gamma \right] - \lambda \left(\frac{c_t - h_{mt}}{(1+r)^t} - I_0 \right) \\ & 0 \leq h_{nt} + h_{mt} \leq H_t \end{aligned}$$

Under this specification, the $H_t - h_{mt} - h_{ht}$ represents the leisure good and the disutility of working manifests itself indirectly in the leisure variable as opposed to the previous model where leisure is indirectly measured as the disutility of working. The labor supply Frisch elasticity for the standard model was 0.106 and the home production model was 0.306. The slightly more complicated home production model had larger estimates for the labor supply elasticities over a large number of different specifications for the standard models and home production models, respectively.

Model

Returning back to our model, consider an optimization problem where the DM simply solves a series of static optimization problems given by

$$\begin{aligned} \max_{\{h_{mt}, L_t | T \geq t \geq 0\}} & \sum_{t=0}^T \beta^t \left\{ \alpha \ln c_t + \frac{(1-\alpha)}{(1-\gamma)} (L_t^{1-\gamma} - 1) \right\} \\ p_t c_t = I_t &= (1-\tau_t) w_t h_{mt} + N_t \\ h_{mt} + L_t &\leq H_t \end{aligned}$$

where N_t is the non-labor income and p_t is the price level for consumption. The FOC of this problem allow us to estimate the parameters of the utility function by running the regression

$$\ln L_t = \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln((1-\tau_t) w_t) + \frac{1}{\gamma} \ln I_t + \varepsilon_t$$

As in Rupert, Rogerson and Wright (2000), we will aggregate the data by age

$$\ln L_{at} = \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln((1-\tau_{at}) w_{at}) + \frac{1}{\gamma} \ln I_{at} + \varepsilon_{at}$$

where the variables have been averaged over the age of each person. The reason for this procedure is that we so not have a true panel data set, so we assume that individual heterogeneity can be removed by averaging over age. In all of the models presented in this chapter, instrumental variables will be applied, the details of which can be found the data section.

As noted above, the elasticity of leisure with respect to after tax wages is not $\frac{-1}{\gamma}$

as one might expect at first glance because income is also a function of after tax wages so that

$$\frac{\partial \ln L_{it}}{\partial \ln(1-\tau_t)w_{it}} = \frac{-1}{\gamma} + \frac{1}{\mathcal{I}_{it}} \left(\frac{\partial I_{it}}{\partial \ln(1-\tau_t)w_{it}} \right)$$

However, we can still calculate the labor supply elasticity with respect to wages, which will result in a very similar expression

$$e_{market}^M = \frac{\partial \ln h_{mt}}{\partial \ln(1-\tau_t)w_t} = \frac{L_t(1-S_t)}{\mathcal{I}_{mt} + L_t S_t} = \frac{(1-S_{mt})S_{Lt}}{S_{mt}(\gamma + S_{Lt})}$$

where $S_{mt} = \frac{(1-\tau_t)w_t h_{mt}}{I_t}$ is the ratio of labor earnings to total income and

$S_{Lt} = \frac{(1-\tau_t)w_t L_t}{I_t}$ is the ratio of virtual labor earnings, which the DM would have

received if their leisure hours were sold in the market place as labor, to full income.

Notice that when $S_{mt} \approx 1$ or $S_{Lt} \approx 0$, then the Marshallian elasticity will be very small. In any given period, it seems reasonable that labor income represents a large proportion of total income, so it should be no surprise that many studies find negligible Marshallian elasticities.

While the wage elasticity is interesting on its own terms, the focus of this paper is on the way labor supply changes with respect to tax changes, not wage changes. However, there is a simple relationship between the two types of elasticities. First note that the after tax wage elasticity is the same as the before tax wage elasticity

$$\frac{\partial \ln h_{mt}}{\partial \ln w_t} = \frac{\partial \ln h_{mt}}{\partial \ln(1-\tau_t)w_t} \frac{\partial(\ln(1-\tau_t) + \ln w_t)}{\partial \ln w_t} = \frac{\partial \ln h_{mt}}{\partial \ln(1-\tau_t)w_t}$$

This allows us to express the tax elasticity in terms of the wage elasticity

$$\frac{\partial \ln h_{mt}}{\partial \ln \tau_t} = \frac{\partial \ln h_{mt}}{\partial \ln(1-\tau_t)w_t} \frac{\partial \ln(1-\tau_t)w_t}{\partial \ln \tau_t} = \frac{\partial \ln h_{mt}}{\partial \ln w_t} \frac{\frac{\partial \ln(1-\tau_t)w_t}{\partial \tau_t}}{\frac{\partial \ln \tau_t}{\partial \tau_t}} = \left(\frac{-\tau_t}{1-\tau_t} \right) \left(\frac{\partial \ln h_{mt}}{\partial \ln w_t} \right)$$

This result simplifies are efforts considerably because we can proceed by deriving the wage elasticity for various models and then multiply by a tax factor that will vary by time and country. However, even if two countries have the same wage elasticity, the tax elasticity will be greater for the country with greater taxes. It is also worth noting that a tax rate of .5 is an interesting watermark to avoid for governments to avoid because

$\frac{\tau_t}{1-\tau_t} \geq 1$ for $\tau_t \geq .5$. This implies that the tax elasticity will be greater than the wage

elasticity whenever the tax rate is greater than .5 and the opposite will be true for tax rates less than .5.

The Marshallian elasticity is not the only measure of labor supply elasticity. The Marshallian elasticity implicitly includes the substitution and income effects. It is also interesting to remove the income effects and focus solely on the substitution effects, which is given by the Hicks elasticity. To derive the Hicks elasticity, recall that we can use the Slutsky equation to write

$$e_{market}^M = \frac{w}{h} \frac{\partial h}{\partial w} = \frac{w}{h} \frac{\partial h}{\partial w} \Big|_u + w \frac{\partial h}{\partial N} = e_{market}^H + w \frac{\partial h}{\partial N}$$

where e_{market}^H is the Hicks elasticity. The Hicks elasticity is a compensated elasticity that accounts for the change in income that occurs with a wage change. We can still appeal to our implicit differentiation approach to write

$$e_{market}^H = e_{market}^M + w \frac{\frac{\partial F_t}{\partial N_t}}{\frac{\partial F_t}{\partial h_{mt}}} = \frac{S_{L_t}}{S_{mt}(\gamma + S_{L_t})} \geq 0$$

A series of static problems optimization problems can be extended to a full dynamic optimization problem with savings by having the DM solve

$$\begin{aligned} \max_{\{c_t, h_{mt}, L_t | 0 \leq t \leq T \leq +\infty\}} & \sum_{t=0}^T \beta^t \left(\alpha \ln(c_t) + \frac{1-\alpha}{1-\gamma} [(L_t)^{1-\gamma} - 1] \right) \\ & (p_t c_t + (1+r)b_{t-1} - (1-\tau_t)w_t h_{mt} - N_t - b_t) \leq 0 \\ & h_{mt} + L_t \leq H_t \end{aligned}$$

where b_t is the net borrowing rate in period t which must be repaid in the next period at the market interest rate, r , and N_t represents non-labor income which can include capital income, transfers, gifts, etc. We assume that the discount factor is normalized so that

$$1 = \sum_{t=0}^T \beta^t$$

This is assumption can be made without any loss of generality because this assumption is

equivalent to rescaling the utility function be a factor of $\frac{1+\beta^{T+1}}{1-\beta}$. Clearly, the entire

time endowment will be used at a maximum, so we can eliminate leisure and form the full Lagrangian, which will be given by

$$V = \sum_{t=0}^T \beta^t \left\{ \alpha \ln c_t + \left(\frac{1-\alpha}{1-\gamma} \right) [(H_t - h_{mt})^{1-\gamma} - 1] \right\} - \lambda_t (p_t c_t + (1+r)b_{t-1} - (1-\tau_t)w_t h_{mt} - N_t - b_t)$$

Examining the FOC for b_t , we find the relation

$$\frac{\partial V}{\partial b_t} = -(1+r)\lambda_{t+1} + \lambda_t = 0$$

This is the classic inter-temporal optimality condition that determines the time path of the Lagrange multiplier, which is given by

$$\lambda_t = \frac{\lambda_0}{(1+r)^t}$$

Assuming an interior solution, the FOC for time t labor supply results in the following relation

$$\ln(H_t - h_{mt}) = \frac{-1}{\gamma} \ln\left(\frac{\lambda_0}{1-\alpha}\right) + \frac{\ln(\beta(1+r))}{\gamma} t - \frac{1}{\gamma} \ln(1-\tau_t) w_t$$

which is the same result as in Rupert, Rogerson and Wright (2000), however, with a little more work we can improve upon this result. The FOC for c_t and c_{t+1} will be given by

$$\begin{aligned} \frac{\partial V}{\partial c_t} &= \frac{\alpha\beta^t}{c_t} - p_t\lambda_t = 0 \\ \frac{\partial V}{\partial c_{t+1}} &= \frac{\alpha\beta^{t+1}}{c_{t+1}} - p_{t+1}\lambda_{t+1} = 0 \end{aligned}$$

Taking the ratio of these two expressions, consumption can be written as

$$p_{t+1}c_{t+1} = \beta(1+r)p_t c_t = \dots = [\beta(1+r)]^{t+1} p_0 c_0$$

which is consistent with a balanced growth path for consumption. If we assume that the solution to this problem is continuously differentiable, then we can solve the problem iteratively by eliminating a subset of the choice variables and solving the new smaller optimization problem.

Since we know that consumption expenditures grows at a constant rate of $\beta(1+r)$ and the langrage multiplier decreases at a constant rate of $\frac{1}{(1+r)}$, then we can eliminate these variables form the optimization problem and solve the new problem

$$\tilde{V} = \sum_{t=0}^T \beta^t \left\{ \alpha \ln \frac{p_0 c_0}{p_t} + \alpha t \ln(\beta + \beta r) + \left(\frac{1-\alpha}{1-\gamma} \right) \left[(H_t - h_{mt})^{1-\gamma} - 1 \right] \right\} - \lambda_0 \left(\beta^t p_0 c_0 - \frac{I_t}{(1+r)^t} \right)$$

where $I_t = (1 - \tau_t)w_t h_{mt} + N_t + b_t - (1+r)b_{t-1}$ is total income at time t including the borrowing position. It is clear from this relationship that we can estimate the parameters, but as mentioned previously λ_0 will be a function of the tax rate and other parameters. However the constraint will now be binding for c_0 , so the FOC for the constraint will be given by

$$\frac{\partial \tilde{V}}{\partial \lambda_0} = \sum_{t=0}^T \beta^t p_0 c_0 - \frac{I_t}{(1+r)^t} = p_0 c_0 - R = 0$$

where $R = \sum_{i=0}^T \frac{I_i}{(1+r)^i}$ is the discounted lifetime wealth of the individual. It should be

pointed out that these are nominal quantities because we included price levels in the original problem and they were washed out of the model, so the inflation rate does not create any labor supply distortions. Our final optimization problem can now be specified without the nuisance constants as

$$\bar{V} = \sum_{t=0}^T \beta^t \left\{ \alpha \ln(R) + \left(\frac{1-\alpha}{1-\gamma} \right) \left[(H_t - h_{mt})^{1-\gamma} - 1 \right] \right\}$$

This allows us to write the FOC for labor supply in terms of lifetime wealth

$$\frac{\partial \bar{V}}{\partial h_{ms}} = \left(\sum_{t=0}^T \beta^t \right) \frac{\alpha}{R} \left[\frac{(1-\tau_s)w_s}{(1+r)^s} \right] - \frac{(1-\alpha)\beta^s}{(H_s - h_{ms})^\gamma} \leq 0$$

It is important to notice that this approach predicts retirement from the labor force. To see this note that at a corner solution

$$\frac{\partial \bar{V}}{\partial h_{ms}} = \frac{\alpha}{R} \left[\frac{(1-\tau_s)w_s}{[\beta(1+r)]^s} \right] - \frac{1-\alpha}{(H_s)^\gamma} < 0$$

If we assume that there is some point during one's career, where wages stop growing or simply grow at a rate that is strictly less than the growth rate of consumption, which is $\beta(1+r)$, then eventually there must be some time period t_r such that

$$\frac{\alpha}{R} \left[\frac{(1-\tau_{t_r})w_{t_r}}{[\beta(1+r)]^{t_r}} \right] - \frac{1-\alpha}{(H_{t_r})^\gamma} < 0$$

Because by assumption $\lim_{t \rightarrow +\infty} \left[\frac{(1-\tau_t)w_t}{[\beta(1+r)]^t} \right] = 0$ and the rest of the terms are time invariant,

so there will be some time period, t_r , where work drops to zero and it will continue to be zero for all periods after that point. It is important to note that this retirement effect will only occur when $\beta(1+r) \geq 1$ because wages can only fall to the minimum wage and $[\beta(1+r)]^t$ will be decreasing in time, so we have implicitly assumed that $\beta(1+r) \geq 1$, but this assumption can be tested. The special case $\gamma = 1$ is quite informative because it allows for a closed form solution of labor supply given by

$$h_{mt} = R \left\{ H_t - \frac{(1-\alpha)[\beta(1+r)]^t}{(1-\tau_t)w_t} \sum_{s=0}^T \frac{(1-\tau_s)w_s H_s H\{h_{ms}\} + N_s}{(1+r)^s} \right\}$$

where R is the ramp function and H is the heavy side step function. This shows the

retirement effect quite clearly as $\frac{[\beta(1+r)]^t}{(1-\tau_t)w_t}$ is increasing in time, so eventually the ramp

function is binding. Another interesting quantity is the summation term in the solution, which we will call

$$\hat{R} = \sum_{s=0}^T \frac{(1-\tau_s)w_s H_s H\{h_{ms}\} + N_s}{(1+r)^s}$$

This relation can be interpreted as the value of a human life. Before one enters the labor force, the value of their life is the discounted value of transfers. The value of life during childhood is simply the discounted value of money the parents spent to raise their child. Likewise in retirement, transfers, including social security and investments, represent the annual value of life. During active working years, the total time endowment is used as the measure of annual value of life in addition to transfers. This concept will be explored further in later sections.

However, at an interior solution we can rearrange the FOC to form regression equation of the form

$$\ln(H_t - h_{mt}) = \frac{-1}{\gamma} \ln\left(\frac{\alpha}{1-\alpha}\right) + \frac{1}{\gamma} \ln R + \frac{\ln \beta(1+r)}{\gamma} t - \frac{1}{\gamma} \ln(1-\tau_t)w_t + \varepsilon_t$$

This equation is similar to the standard labor supply estimation model except for the lifetime, discounted wealth, which has replaced λ_0 . Under the standard approach,

$$\frac{\partial \ln(H_t - h_{mt})}{\partial \ln(1-\tau_t)w_t} = \frac{\partial \ln \lambda_0}{\partial \ln(1-\tau_t)w_t} + \frac{-1}{\gamma} = \frac{-1}{\gamma}$$

However, it is not very reasonable to assume λ_0 is exogenous, but we can now quantify the error of this assumption because

$$\frac{\partial \ln(H_t - h_{mt})}{\partial \ln(1 - \tau_t)w_t} = \frac{-1}{\gamma} (1 - \bar{S}_{mt}) > \frac{-1}{\gamma}$$

where $\bar{S}_{mt} = \frac{(1 - \tau_t)w_t h_{mt}}{(1 + r)^t R}$ is the ratio of discounted labor income in period t to

discounted lifetime income. It is interesting to note that under the assumption that income grows at the risk free interest rate, then

$$\bar{S}_{mt} = \frac{(1 - \tau_t)w_t h_{mt}}{(1 + r)^t T I_o}$$

which means the approximation $\bar{S}_{mt} \approx 0$ will be more appropriate in long-term optimization problems and it will hold exactly in the infinite horizon problem, but constant income growth might be too restrictive in many cases. If we simplify the regression equation further, then a rather interesting result arises.

$$\ln(H_t - h_{mt}) = \frac{-1}{\gamma} \ln\left(\frac{\alpha}{1 - \alpha}\right) + \frac{t}{\gamma} \ln \beta + \frac{1}{\gamma} \ln[(1 + r)^t R] - \frac{1}{\gamma} \ln(1 - \tau_t)w_t + \varepsilon_t$$

In perfect markets without uncertainty, the DM chooses their net borrowing positions so that the life-cycle paths have the appearance of the DM receiving their entire lifetime income, which will be discounted, and then the DM receives a stream of income growing at the rate $\beta(1 + r)$. While it may be difficult to measure discounted lifetime wealth from an econometric aspect, we can still get viable estimates by first differencing the regression equation

$$\Delta \ln(H_t - h_{mt}) = \frac{1}{\gamma} \ln \beta(1 + r) - \frac{1}{\gamma} \Delta \ln(1 - \tau_t)w_t + \varepsilon_t$$

where $\Delta \ln(H_t - h_{mt}) = \ln(H_t - h_{mt}) - \ln(H_{t-1} - h_{mt-1})$ which allows the interest rate, r , and the intertemporal substitution parameter, γ , to be estimated. This formulation also shows that the trajectory of leisure will be determined by the trajectory of wages. Using the implicit differentiation approach the labor supply elasticity with respect to wages will be

$$\bar{e}_{market}^M = \frac{\partial \ln h_{mt}}{\partial \ln(1 - \tau_t)w_t} = \frac{(1 - \bar{S}_{mt})\bar{S}_{Lt}}{\bar{S}_{mt}(\gamma + \bar{S}_{Lt})}$$

This expression is the same as before except \bar{S}_{mt} has been replaced by S_{mt} and \bar{S}_{Lt} has replaced S_{mt} , so now we are taking the ratio of labor income to total income in period t as opposed to taking the ratio of discounted labor income in time t to discounted lifetime income.

Since we have labor supply elasticity for each individual, the average elasticity can be computed over all individuals, which gives a reasonable estimate of aggregate labor elasticity, which allows for hypothesis testing to be computed at each point in time. While this method doesn't allow elasticity to be computed as an explicit function of time, it does allow us to see how labor elasticity has evolved over time as the tax rate has changed.

In the previous model, the DM must substitute market labor for leisure whenever taxes change. However, this approach ignores the fact that people provide labor outside the market place that is fundamentally different from leisure. Whenever one cooks a meal at home, they are a part of the production process of the meal, which is cheaper, and doesn't face the same tax burden as eating at a restaurant. In a similar fashion, as in the

simulation experiments, it is assumed that the DM solves a series of static optimization problems

$$\begin{aligned} \max_{\{I_t, h_{mt}, h_{ht}, L_t | t \geq 0\}} \sum_{t=0}^{\infty} \beta^t & \left\{ \frac{\alpha}{\varepsilon} \ln [\theta c_t^\varepsilon + (1-\theta)h_{ht}^\varepsilon] + \frac{(1-\alpha)}{(1-\gamma)} (L_t^{1-\gamma} - 1) \right\} \\ p_t c_t = I_t &= (1-\tau_1)w_t h_{mt} + N_t \\ h_{mt} + h_{ht} + L_t &\leq H_t \end{aligned}$$

The FOC are given by

$$\begin{aligned} \frac{\partial L}{\partial h_{mt}} &= \frac{\beta^t \alpha \theta (1-\tau_1) \tilde{w}_t \tilde{I}_t^{\varepsilon-1}}{\tilde{\theta} \tilde{I}_t^\varepsilon + (1-\theta)h_{ht}^\varepsilon} - \lambda = 0 \\ \frac{\partial L}{\partial h_{ht}} &= \frac{\beta^t \alpha (1-\theta)h_{ht}^{\varepsilon-1}}{\tilde{\theta} \tilde{I}_t^\varepsilon + (1-\theta)h_{ht}^\varepsilon} - \lambda = 0 \\ \frac{\partial L}{\partial L_t} &= \frac{(1-\alpha)\beta^t}{L_t^\gamma} - \lambda = 0 \end{aligned}$$

Where $\tilde{I}_t = \frac{I_t}{p_t}$ and $\tilde{w}_t = \frac{w_t}{p_t}$. From these FOC, it can be deduced that

$$\begin{aligned} h_{ht} &= \left[\frac{\theta(1-\tau_1)w_t}{(1-\theta)p_t} \right]^{\frac{1}{\varepsilon-1}} \tilde{I}_t = A_t \tilde{I}_t \\ L_t &= \left[\left(\frac{1-\alpha}{\alpha} \right) \left(\frac{1+(1-\tau_1)\tilde{w}_t A_t}{(1-\tau_1)\tilde{w}_t} \right) \tilde{I}_t \right]^{1/\gamma} \end{aligned}$$

These two relations will hold for all individuals. These relations will change over time, but they change only in the sense that the taxes, wages, and transfers change over time, so the parameters, which are time invariant, can be estimated from the following system of equations

$$\begin{aligned} \ln h_{hit} &= \frac{1}{\varepsilon-1} \ln \left(\frac{\theta}{1-\theta} \right) + \frac{1}{\varepsilon-1} \ln ((1-\tau_1)w_{it}) + \ln I_{it} + \frac{\varepsilon}{1-\varepsilon} \ln p_t + \mu_{it} \\ \ln L_{it} &= \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln ((1-\tau_1)\tilde{w}_{it}) + \frac{1}{\gamma} \ln \tilde{I}_{it} + \frac{1}{\gamma} \ln (1 + (1-\tau_1)\tilde{w}_{it} A_{it}) + \nu_{it} \end{aligned}$$

It is interesting to note that by using the relation $h_{hit} = A_{it} \tilde{I}_{it}$ the second equation can be rewritten giving

$$\begin{aligned}\ln h_{hit} &= \frac{-1}{1-\varepsilon} \ln\left(\frac{\theta}{1-\theta}\right) - \frac{1}{1-\varepsilon} \ln((1-\tau_t)w_{it}) + \ln I_{it} + \frac{\varepsilon}{1-\varepsilon} \ln p_t + \mu_{it} \\ \ln L_{it} &= \frac{-1}{\gamma} \ln\left(\frac{\alpha}{1-\alpha}\right) - \frac{1}{\gamma} \ln((1-\tau_t)w_{it}) + \frac{1}{\gamma} \ln(I_{it} + (1-\tau_t)w_{it}h_{hit}) + \nu_{it}\end{aligned}$$

The third term on the right side of the second equation gives the hypothetical income the DM would have had they sold their home production labor in the open market. If we assume $\text{cov}(\mu_{it}, \nu_{it}) = 0$, then the system of equations will be fully recursive, which means that OLS will be consistent and efficient. As with the non-home production model, leisure is not directly affected by the price level; however, home production is affected. Taking first differences

$$\begin{aligned}\Delta \ln h_{hit} &= \Delta \ln I_{it} - \frac{1}{1-\varepsilon} \Delta \ln((1-\tau_t)w_{it}) + \frac{\varepsilon}{1-\varepsilon} \Delta \ln p_t + \Delta \mu_{it} \\ \Delta \ln L_{it} &= -\frac{1}{\gamma} \Delta \ln((1-\tau_t)w_{it}) + \frac{1}{\gamma} \Delta \ln(I_{it} + (1-\tau_t)w_{it}h_{hit}) + \Delta \nu_{it}\end{aligned}$$

Consider the scenario where the consumption price level is increasing while holding wages fixed with $\varepsilon > 0$. It clearly cannot be the case that both home production and market labor supply are increasing because our leisure equation is increasing in income and home production and all three choice variables cannot be increasing. Similar reasoning indicates that home production and market labor supply cannot decrease at the same time. Under the assumption that $\varepsilon > 0$, home production is increasing with the price level which means market labor supply is decreasing. This assumption also implies that the market consumption and home production are elastic, so it is easy for people to evade the market by producing at home. As people cut back on their market labor

supply, by moving from full time to part time or cutting back full time hours, or even leaving the job market, the equilibrium wage will rise. Remembering that the unemployment rate is essentially labor demand shortage whereby the unemployed cannot find buyers for labor supply. Firms will start to hire more people and the unemployment rate goes down. However $\varepsilon < 0$ requires that home production is decreasing in a high inflation environment because home production and market consumption are inelastic, so workers cannot evade inflation as much by leaving the market which means workers are more willing to delay retirement, become full time workers, or work harder. This allows firms to cut jobs resulting in higher unemployment rates, which is stagflation. We will pursue this issue further in the dynamic setting where we show that the sign of the labor supply elasticity with respect to consumption prices is entirely dependent on the elasticity between home production and market consumption.

The estimates of the parameters of interest are not varying over time or across individuals, so for any given individual the labor supply elasticity at time t can be found by implicit differentiation giving

$$e_{market}^{MH} = \frac{\partial \ln h_{mt}}{\partial \ln(1 - \tau_t)w_t} = \frac{(S_{Lt} + \gamma S_{ht})[1 - (1 - \varepsilon)S_{mt}] - \varepsilon \left(1 - \frac{(1 - \tau_t)w_t}{z_t}\right) S_{Lt}}{(1 - \varepsilon)S_{mt}(\gamma + S_{Lt} + \gamma S_{ht})}$$

where $z_t = \left(\frac{\theta}{1 - \theta}\right)^{\frac{1}{1 - \varepsilon}} [(1 - \tau_t)w_t]^{\frac{1}{1 - \varepsilon}} p_t^{\frac{-\varepsilon}{1 - \varepsilon}} + (1 - \tau_t)w_t$ and e_{market}^{MH} is the Marshallian labor

supply elasticity in the home production model. Benhabib, J., Richard Rogerson, and Randall Wright (1991) showed that in the home production model whenever $\varepsilon = 0$ and $\gamma = 1$, home production has no effect on market labor supply decisions, which is to say

home production can be ignored. First note that in the home production model, leisure includes home production indirectly, which means

$$S_{L_t} + S_{h_t} = \frac{(1 - \tau_t)w_t(H_t - h_{mt})}{I_t}$$

which is exactly our expression for S_{L_t} in the non-home production model. Finally

evaluating at $\varepsilon = 0$ and $\gamma = 1$, we find that

$$e_{market}^{MH} = \frac{\left[\frac{(1 - \tau_t)w_t(H - h_{mt})}{I_t} \right] [1 - S_{mt}]}{S_{mt} \left(1 + \left[\frac{(1 - \tau_t)w_t(H - h_{mt})}{I_t} \right] \right)} = e_{market}^M$$

So we find that under these parameter values, home production can be safely ignored, but more importantly the data can be used to test the validity of including home production in the labor supply model.

The sign the elasticity in the home production model is ambiguous unlike in the non-home production, which has a strictly positive Marshallian elasticity. Since some economic studies have found evidence of a negative Marshallian wage elasticity, it is important that our model allows for such a possibility. Under the non-home production model, we cannot make a meaningful rejection of the hypothesis that the Marshallian wage elasticity is negative because it is positive by construction, but the home production model will allow us to do just that.

As with the non-home production model, we can use the same implicit differentiation approach to calculate the Hicks elasticity, which will be given by

$$e_{market}^{HH} = \frac{(S_{Lt} + \gamma S_{ht}) - \varepsilon \left(1 - \frac{(1 - \tau_t)w_t}{z_t} \right) S_{Lt}}{(1 - \varepsilon) S_{mt} (\gamma + S_{Lt} + \gamma S_{ht})}$$

If the Hicks elasticity is evaluated at $\varepsilon = 0$ and $\gamma = 1$, we would again find that the home production model does not affect the labor supply decision problem.

While the Marshallian elasticity can be negative, the Hicks elasticity must be positive, which we can verify by rewriting the Hicks elasticity as

$$e_{market}^{HH} = \frac{\gamma S_{ht} + \left[1 - \varepsilon \left(1 - \frac{(1 - \tau_t)w_t}{z_t} \right) \right] S_{Lt}}{(1 - \varepsilon) S_{mt} (\gamma + S_{Lt} + \gamma S_{ht})} \geq 0$$

Since $\varepsilon \leq 1$ and $\left(1 - \frac{(1 - \tau_t)w_t}{z_t} \right) \leq 1$, the Hicks elasticity must be positive reinforcing the economic theory of our model.

The same techniques that were used in the non-home production model can be used in the home production model to extend the series of static problems to a fully dynamic setting with savings. The dynamic optimization will be given by

$$\begin{aligned} \max_{\{c_t, h_{mt}, h_{ht}, L_t \mid 0 \leq t \leq T \leq +\infty\}} & \sum_{t=0}^T \beta^t \left(\frac{\alpha}{\varepsilon} \ln(c_t^\varepsilon + (1 - \theta)h_{ht}^\varepsilon) + \frac{1 - \alpha}{1 - \gamma} [(L_t)^{1 - \gamma} - 1] \right) \\ & (p_t c_t + (1 + r)b_{t-1} - (1 - \tau_t)w_t h_{mt} - N_t - b_t) \leq 0 \\ & h_{mt} + h_{ht} + L_t \leq H_t \end{aligned}$$

In the previous model, it was assumed consumption and labor at time t were chosen based solely on the income from that period. It will be shown that the fully dynamic model is essentially the same except the notion of income has changed from period t income to lifetime, discounted income. First subsume the borrowing decision into income and write the full Lagrangian as

$$\bar{V} = \sum_{t=0}^T \beta^t \left\{ \frac{\alpha}{\varepsilon} \ln(\theta c_t^\varepsilon + (1-\varepsilon)h_{ht}^\varepsilon) + \left(\frac{1-\alpha}{1-\gamma} \right) [L_t^{1-\gamma} - 1] \right\} - \frac{\lambda_0}{(1+r)^t} (p_t c_t - I_t)$$

The FOC will be given by

$$\begin{aligned} h_{mt} \frac{\partial \bar{V}}{\partial h_{mt}} &= h_{mt} \left[\frac{-\beta^t (1-\alpha)}{L_t^\gamma} + \frac{(1-\tau_t)w_t \lambda_0}{(1+r)^t} \right] = 0 \\ h_{ht} \frac{\partial \bar{V}}{\partial h_{ht}} &= \left[\frac{\alpha \beta^t (1-\theta) h_{ht}^\varepsilon}{\theta c_t^\varepsilon + (1-\theta) h_{ht}^\varepsilon} - \frac{(1-\alpha) \beta^t h_{ht}}{L_t^\gamma} \right] = \left[\frac{\alpha \beta^t (1-\theta) h_{ht}^\varepsilon}{\theta c_t^\varepsilon + (1-\theta) h_{ht}^\varepsilon} + \frac{(1-\tau_t)w_t \lambda_0 h_{ht}}{(1+r)^t} \right] = 0 \\ c_t \frac{\partial \bar{V}}{\partial c_t} &= \left[\frac{\alpha \beta^t (\theta) c_t^\varepsilon}{\theta c_t^\varepsilon + (1-\theta) h_{ht}^\varepsilon} - \frac{\lambda_0 p_t c_t}{(1+r)^t} \right] = 0 \end{aligned}$$

where the first equation was substituted into the second equation. If the second equation and third equations are added together, the result is the following relation.

$$p_t c_t + (1-\tau_t)w_t h_{ht} = z_t h_{ht} = \frac{\alpha [\beta(1+r)]^t}{\lambda_0}$$

where z_t was previously defined. In the dynamic non-home production mode, consumption expenditures were growing at a constant rate, which is the same rate as in the home production model. The main difference is that the DM acts as if home production is a consumption good being bought at the after tax wage rate. From the second and third equations, it can be deduced that

$$h_{ht} = \left[\frac{\theta(1-\tau_t)w_t}{(1-\theta)p_t} \right]^{\frac{1}{\varepsilon-1}} c_t = A_t c_t$$

which is equivalent to the expression for home production that was previously derived except income has been replaced by consumption. These two relations allow the trajectory of home production to be written as

$$z_t h_{ht} = [\beta(1+r)] z_{t-1} h_{ht-1} = \dots = [\beta(1+r)]^t z_0 h_0$$

These relations allow consumption to be eliminated and the new Lagrangian will be

$$\bar{V}_2 = \sum_{t=0}^T \beta^t \left\{ \frac{\alpha}{\varepsilon} \ln(\theta A_t^{-\varepsilon} + (1-\theta)) + \alpha \ln h_{ht} + \left(\frac{1-\alpha}{1-\gamma} \right) [L_t^{1-\gamma} - 1] \right\} - \frac{\lambda_0}{(1+r)^t} (z_t h_{ht} - \hat{I}_t)$$

where $\hat{I}_t = I_t + (1-\tau_t)w_t h_{ht}$ is the virtual income that would have been available if home production was sold in the marketplace as labor. In this new problem, the budget constraint can be eliminated

$$\frac{\partial \bar{V}_2}{\partial \lambda_0} = \sum_{t=0}^T - \left(\beta^t z_0 h_{h0} - \frac{\hat{I}_t}{(1+r)^t} \right) = \hat{R} - z_0 h_{h0} = 0$$

where $\hat{R} = \sum_{t=0}^T \frac{I_t + (1-\tau_t)w_t h_{ht}}{(1+r)^t}$ is the discounted, lifetime, virtual income. This relation determines the initial conditions of the dynamic optimization, so the optimal control policy is given by

$$z_t h_{ht} = [\beta(1+r)]^t \hat{R}$$

We would like to get home production in terms of the discounted lifetime income, which can be accomplished by rewriting the expression as

$$z_t h_{ht} = [\beta(1+r)]^t \hat{R} = [\beta(1+r)]^t \left(1 + \sum_{i=1}^T \bar{S}_{hi} \right) R$$

where $\bar{S}_{hi} = \frac{(1-\tau_i)w_i h_{hi}}{(1+r)^i R}$ is the ratio of labor earnings to discounted lifetime income.

This relation can be used to estimate the home production parameters. First rewrite the home production relation

$$(1-\tau_t)w_t \left\{ \beta^t \left(1 + \sum_{i=0}^T \bar{S}_{hit} \right) - \bar{S}_{hat} \right\} = \frac{(z_t - (1-\tau_t)w_t)(1-\tau_t)w_t h_{ht}}{(1+r)^t R}$$

Taking the logarithm of both sides and rearranging, we arrive at the regression equation

$$\ln \bar{S}_{hat} - \ln \left[\beta^t \left(1 + \sum_{j=a}^T \bar{S}_{hjt} \right) - \bar{S}_{hat} \right] = \frac{-1}{1-\varepsilon} \ln \frac{\theta}{1-\theta} - \frac{\varepsilon}{1-\varepsilon} \ln \frac{(1-\tau_t)w_{at}}{p_t} + v_{at}$$

This equation allows ε and θ to be identified, which constitute all of the home production parameters.

While a constant interest rate does simplify the model, a time varying interest rate can be incorporated into the model relatively easily. If a time varying interest rate is used, everything in the model will be the same except $\prod_{i=1}^t (1+r_i)$ replaces $(1+r)^t$ throughout.

So under this specification

$$\hat{R}_t = \sum_{i=0}^T \frac{I_{ai} + (1-\tau_t)w_{ai}h_{hai}}{\prod_{j=0}^i (1+r_j)}$$

$$\ln(z_t h_{ht}) = t \ln \beta + \sum_{i=1}^t \ln(1+r_i)$$

This would seem to make more sense, but if we have a 1990 survey and our oldest cohort is 68, then we would need interest rates from 1940 to form this lifetime discounted income because 1940 is the year our 68 year old cohorts turned 18. For example 1975 is our earliest survey for the US, so we would need interest rate data from 1928, however, 1954 is the earliest time period we could get for US interest rates from the OECD database. So in our analysis we use a time varying interest rate and replacing any missing interest rates with the one that is as close as possible in time. Using the example above, the 1954 interest rate would be used for the first 14 periods and the time varying interest rate would be used for the final 36 periods.

In order to find the labor supply elasticity, we need to form an implicit function of the wage and market labor supply, which means home production must be eliminated. Our recurrence relation for home production still depends on the sequence of home production choices, but after a bit of manipulation, it can be shown that

$$\frac{\hat{R}}{R} = 1 + \sum_{i=0}^T \bar{S}_{hi} = \frac{1}{\sum_{i=0}^T \frac{\beta^i p_i A_i^{-1}}{z_i}} = \frac{1}{D}$$

This verifies that the optimal control for home production can be written in terms of the market labor supply decisions made over the life-cycle of the decision maker. This allows the budget constraint to be eliminated directly to produce the final Lagrangian

$$\bar{V}_3 = \sum_{i=0}^T \beta^i \left\{ \frac{\alpha}{\varepsilon} \ln(\theta A_i^{-\varepsilon} + (1-\theta)) + \alpha \ln \left[\frac{g_i}{z_i} \right] + \alpha \ln \hat{R} + \left(\frac{1-\alpha}{1-\gamma} \right) [L_i^{1-\gamma} - 1] \right\}$$

The problem has been reduced to a function of the total work, which is to say the sum of home production and market labor, so the FOC will be given by

$$\frac{\partial \bar{V}_3}{\partial h_{ms}} = \frac{\partial \bar{V}_3}{\partial h_{hs}} = \left(\sum_{i=0}^T \beta^i \right) \left[\frac{\alpha(1-\tau_s)w_s}{(1+r)^s \hat{R}} \right] - \frac{(1-\alpha)\beta^s}{L_s^\gamma} = 0$$

From this relation, the optimal control policy for leisure will be

$$L_t = \left[\frac{(1-\alpha)[\beta(1+r)]^t \hat{R}}{\alpha(1-\tau_t)w_t} \right]^{\frac{1}{\gamma}}$$

Using this relation, we can estimate γ and $\beta(1+r)$ by running the regression

$$\ln L_{it} = \frac{-1}{\gamma} \ln \frac{\alpha}{1-\alpha} - \frac{1}{\gamma} \ln(1-\tau_t)w_{it} + \frac{t}{\gamma} \ln \beta(1+r) + \frac{1}{\gamma} \ln \hat{R}_i + \varepsilon_{it}$$

Since our data set is cross-sectional, \hat{R}_i is not computable at the individual level, but we can aggregate the data by age which will allow us to calculate the discounted, lifetime, virtual income.

All other things being equal, we can see that leisure will not grow at the same rate as home production. Nevertheless the time constraint can be used to deduce a relation for market labor supply

$$h_{mt} = H_t - h_{ht} - L_t = R \left\{ H_t - \frac{[\beta(1+r)]^t \hat{R}}{z_t} - \left[\frac{(1-\alpha)[\beta(1+r)]^t \hat{R}}{\alpha w_t} \right]^{\frac{1}{\gamma}} \right\}$$

where $R\{x\}$ is the ramp function. This shows the retirement effect discussed earlier. If leisure or home production increases over time, eventually market labor drops to zero and the decision maker retires. If market labor is positive, there is a certain minimum of hours that need to be worked in order to make it a worthwhile for both parties. This minimum is hasn't been included in the model, but this issue does have distortionary effect in the market. For example childrearing is a very time intensive home production activity especially when the children are young. The large increase in home production demands can drive people out of the labor market because they need to decrease their labor supply below the minimum requirement. Of course, women are disproportionately affected by this phenomenon. If we are considering a minimum of hours, h_{mt}^{\min} , that must be met in order to stay employed, there is a differentiation in unemployment.

Suppose a worker is earning a wage that is below the market rate for their skill set. In this scenario, we can model this scenario as a worker having a permanent wage cut. At this new wage the worker would like to decrease labor supply, but the minimum

number of hours effect may force the worker to look for a job with better pay or hours. This can be considered frictional unemployment because the worker would remain unemployed if the old job was available, provided the expected loss of wages during the job search is less than the expected gains of a better job. However, if the worker is laid off by the firm, then the worker would be utility maximizing under the old job conditions which allows us to write unemployment as

$$U_t = \frac{\sum_{i=1}^N H \{h_{m(t-1)}^{\min} - h_{imt}^S\}}{\sum_{i=1}^N H \{h_{imt}^S\}} + \frac{\sum_{i=1}^N (1 - H \{h_{imt}\}) H \{h_{imt}^S\}}{\sum_{i=1}^N H \{h_{imt}^S\}}$$

where h_{imt}^S is the quantity of labor that the worker would like to supply given the current parameters. The first term is the frictional unemployment and the second term is structural unemployment. The second term is of more interest from an economic perspective, but this approach does provide an intuition on how to separate these two effects.

The natural unemployment can be computed simply as the proportion of the unemployed that would accept the same job at a different firm in a similar location. The unemployment survey could be altered to ask if the respondent would accept the exact same job at a different firm. If the answer is in the affirmative, then the respondent is structurally unemployed. If the answer is in the negative, then worker is searching for better opportunities and should not be included in the natural rate of unemployment. These quantities can be estimated and our general approach could be generalized to include unemployment in future research.

While these relations for leisure and home production are not closed form solutions, they do provide a functional form that can be estimated with standard econometric tools. After applying the method of implicit differentiation, we arrive at the Marshallian labor supply elasticity for the home production model.

$$\bar{e}_{market}^{MH} = \frac{(\bar{S}_{Lt} + \gamma \bar{S}_{ht}) \left[1 - \varepsilon \left(\frac{(1 - \tau_t) w_t}{z_t} \right) \left(1 + \bar{S}_{ht} - \beta^t \left(1 + \sum_{i=1}^T \bar{S}_{hi} \right) \right) - (1 - \varepsilon) \bar{S}_{mt} \right] - \varepsilon \bar{S}_{Lt} \left(1 - \frac{(1 - \tau_t) w_t}{z_t} \right)}{(1 - \varepsilon) \bar{S}_{mt} (\gamma + \bar{S}_{Lt} + \gamma \bar{S}_{ht})}$$

While this expression can be negative, the wage elasticity will be positive for virtually all of our results. As discussed earlier, the change in labor supply with respect to consumption prices is intimately tied to the elasticity between home production and market consumption.

$$\frac{\partial h_{mt}}{\partial p_t} = \left(\frac{-\varepsilon}{1 - \varepsilon} \right) \left(\frac{A_t}{z_t} \right) \left[\frac{\mathcal{H}_{ht} + \left(\frac{\beta^t (1 - \tau_t) w_t}{D z_t} \right) (\mathcal{H}_{ht} + L_t)}{\gamma + \gamma \bar{S}_{Lt} + \gamma \bar{S}_{ht}} \right]$$

Here it is much more clear that labor supply and prices will move in the same direction whenever $\varepsilon < 0$. Correspondingly the labor supply will move in the opposite direction of prices. The surprising result is that when $\varepsilon = 0$ market labor supply is unaffected by price levels. This clearly shows the distortionary affects created by inflation whenever home production is relevant. Furthermore It can be easily verified that whenever $\varepsilon = 0$ and $\gamma = 1$, the home production elasticity will coincide with the non-home production elasticity in the fully dynamic model.

In order to eliminate income effects, the Hicks wage elasticity can be computed as

$$\bar{e}_{market}^{HH} = \frac{(\bar{S}_{Lt} + \gamma \bar{S}_{ht}) \left[1 - \varepsilon \left(\frac{(1 - \tau_t) w_t}{z_t} \right) \left(1 + \bar{S}_{ht} - \beta^t \left(1 + \sum_{i=1}^T \bar{S}_{hi} \right) \right) \right] - \varepsilon \bar{S}_{Lt} \left(1 - \frac{(1 - \tau_t) w_t}{z_t} \right)}{(1 - \varepsilon) \bar{S}_{mt} (\gamma + \bar{S}_{Lt} + \gamma \bar{S}_{ht})} \geq 0$$

While this expression may appear to have an ambiguous sign, it is a simple exercise to verify that the Hicks elasticity will always be positive. Examining the differences between the static and dynamic problems, it becomes apparent that the expressions are the same except the notions of income and savings are on different time scales. In general the dynamic models view labor decisions through the lens income over the lifetime; meanwhile, the static models view labor decisions through immediate market conditions.

Estimation

The majority of the data was compiled from the Multinational Time Use Study. Since this was an international study, there are issues with comparisons across countries, but this study has been harmonized so that we can make meaningful comparisons across studies. Unfortunately, the harmonization creates a sparse data set as many categories we create are not available in all of the countries. The wage and income data was very sparse, which forced us to ignore a large number of countries because wage or income data was not a part of their survey. All of the surveys are cross sectional. Even within surveys, there is a low response rate to wage and income questions. While there is certainly a sample selection problem in those respondents that choose not to report income data, we do not address this issue because we were able to reproduce similar

results using other models that have been used in the literature, which was outlined previously in the literature review. We also compared the descriptive statistics of the variables we used in our analyses with those from more complete national surveys and found that they were quite similar.

In the table below, different time use activities are aggregated into the relevant categories for our model.

Market Labor	Home Production	Leisure	Basic
Paid work	Cook	Gardening	Sleep
Paid work at home	Housework	travel	Dress
Second job	Odd jobs	Excursions	Personal care
School/classes	Childcare	Sports	Shopping
Travel to/from work	Domestic travel	Walking	Meals/snacks
Homework/study	Entertain friends at home	Religious activities	
	Unpaid work at home	Civic activities	
		Cinema	
		Theatre	
		Dance	
		Parties	

		Social clubs	
		Restaurant	
		Listen to Audio	
		Watch T.V.	
		Read	
		Relax	
		Conversation	
		Visit friends at their homes	
		Other leisure	

Clearly market labor, home production, and leisure cannot be the only activities that can be pursued. The basic variables are subtracted from the time endowment and they represent activities that must be done to survive. Gardening might be considered home production, but it doesn't include farming. Any farming activities are to be coded as "paid work at home" if the goods are sold. "Meals/Snacks" was put into the basic variable because the labor codes didn't include eating as part of work. However, it is highly likely that there is a bit of bias introduced with this methodology, although we are less interested in the levels of time use. The exchange and consumer price levels were all downloaded from the OECD website in 2008 US dollars. The interest rate data is also from the OECD website and it is the annualized 10 year bond rate. In cases where interest rate, exchange rate, or consumption price level was unavailable, we imputed the

missing data by using the available data that was closest in time. The instrumental variables used in the regressions of all of the models are given below.

Instrumental Variable	Type
Day	Integer
Month	Integer
Household size	Integer
Number of children	Integer
Age of youngest child	Integer
Home ownership	Indicator
Urban area	Indicator
Computer ownership	Indicator
Vehicle ownership	Indicator
Sex	Indicator
Age	Integer
Living in parent's home	Indicator
Single parent household	Indicator
Citizen Status	Indicator
Employment status	Indicators
Unemployed	Indicator
Student	Indicator
Retirement status	Indicator
Employment status	Indicators
Occupation	Indicators
Sector	Indicators
Education	Indicators

These variables were used whenever they were available to be used. Some countries had missing data for some of the instruments. The time variables were included to adjust for seasonality and weekends. Variables relating to children and the home affect home production. The rest of the variables relate to wages and labor. Since the data is aggregated over age, the aggregated indicators become probabilities for different age groups.

The 1991-1992 Germany data recorded age in five-year bands. In order to deal with this problem, a simple change of time scale was made so that labor was measured over a five-year period. Income was assumed to be constant over five years and was discounted back accordingly. Unfortunately, only 11 data points remain after this procedure as opposed to 51 data points for the other countries. Unsurprisingly, the parameter estimates are not significant at any reasonable level, but the results are included.

The next step in the process was to create a synthetic cohort. All workers in our data set were chosen to be between the ages of 18 and 68 and they must have reported positive market labor, home labor, and leisure so that we could ensure an interior solution to the maximization problem. This restriction could be relaxed in future work. In order to create a synthetic cohort, we average the variables for all people of the same age: for example,

$$L_{at} = \frac{1}{N_a} \sum_{i=1}^{N_a} L_{ait}$$

$$\ln L_{at} = \frac{1}{N_a} \sum_{i=1}^{N_a} \ln L_{ait}$$

where N_a is the number people of age a . For the static model without home production, the following regression will allow us to recover the necessary parameters for the elasticities.

$$\ln L_{at} = \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln(1-\tau_t) w_{at} + \frac{1}{\gamma} \ln I_{at} + \varepsilon_{at}$$

Our hourly wage measure is monthly labor income divided by monthly hours worked, so our hourly wage variable is endogenous by design, so we proceed by using

the instrumental variables outlined above. The dynamic home production model without home production is the same except for the income variable. Since it is a simple generalization to estimate the model with a floating interest rate, then we will run the slightly modified model

$$\ln L_{at} = \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln(1-\tau_t) w_{at} + \frac{1}{\gamma} \left[\ln R + \sum_{t=1}^a \ln(1+r_t) \right] + \frac{\ln \beta}{\gamma} a + \varepsilon_{at}$$

The regression will be performed over the 50 years of age data that we have from the 18 to 68 year olds. The static model with home production will have a system of equations that must be estimated via 2SLS.

$$\begin{aligned} \ln h_{hit} &= \frac{-1}{1-\varepsilon} \ln \left(\frac{\theta}{1-\theta} \right) - \frac{1}{1-\varepsilon} \ln((1-\tau_t) \tilde{w}_{it}) + \ln \tilde{I}_{at} + \mu_{at} \\ \ln L_{it} &= \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln((1-\tau_t) \tilde{w}_{it}) + \frac{1}{\gamma} \ln(\tilde{I}_{it} + (1-\tau_t) \tilde{w}_{it} h_{hit}) + \nu_{at} \end{aligned}$$

where the wage and income variables are have been adjusted for inflation and currency to be 2008 US dollars. The dynamic model will also have a similar system of equations that are to be estimated.

$$\begin{aligned} \ln \bar{S}_{hat} - \ln \left[\beta^t \left(1 + \sum_{j=1}^T \bar{S}_{hjt} \right) - \bar{S}_{hat} \right] &= \frac{-1}{1-\varepsilon} \ln \frac{\theta}{1-\theta} - \frac{\varepsilon}{1-\varepsilon} \ln(1-\tau_t) \tilde{w}_{it} + \mu_{at} \\ \ln L_{at} &= \frac{-1}{\gamma} \ln \left(\frac{\alpha}{1-\alpha} \right) - \frac{1}{\gamma} \ln(1-\tau_t) w_{at} + \frac{a}{\gamma} \ln \beta + \frac{1}{\gamma} \left[\ln \left(1 + \sum_{j=1}^T \bar{S}_{hjt} \right) R + \sum_{t=1}^a \ln(1+r_t) \right] + \nu_{at} \end{aligned}$$

Finally the system of equations are estimated with 2SLS with the instrumental variables mentioned previously.

Results

In the first stage of our analysis of the data, we examined the role of lifetime, discounted income, R , in the model. Recall that

$$R_t = \sum_{i=0}^T \frac{I_i}{\prod_{j=0}^i (1+r_j)}$$

$$\hat{R}_t = \sum_{i=0}^T \frac{I_i + (1-\tau_t)w_i h_{hi}}{\prod_{j=0}^i (1+r_j)}$$

So R is the measure of market labor, investments, and transfers discounted back to the initial time period. Meanwhile \hat{R} is the value of market income plus the amount of money that would have been earned provided home labor was sold in the market place. The next step is to add leisure economic value of leisure if it was sold in the market, which gives

$$\hat{\hat{R}}_t = \sum_{i=0}^T \frac{I_i + (1-\tau_t)w_i (h_{hi} + L_i)}{\prod_{j=0}^i (1+r_j)}$$

This is interpreted as the economic value the individual places on their working life, namely the ages between 18 and 68. In the table below, the various measures are reported in millions of 2008 US dollars so that international and intertemporal comparisons can be made.

Country	Year	R	\hat{R}	$\hat{\hat{R}}$
Canada	1981	4.5492	5.7618	7.4122
Canada	1986	4.9806	6.2774	8.0395

Norway	2000	3.5298	4.0862	4.9166
Germany	1991	4.4136	4.9715	5.6206
Germany	1992	4.3656	4.8879	5.5094
Germany	2001	3.2977	3.7387	4.3040
Germany	2002	3.3504	3.7676	4.3017
Spain	2002	5.1837	5.9392	7.2920
Spain	2003	6.3426	7.2210	8.9272
UK	2000	8.5950	10.0537	12.3005
UK	2001	7.7922	9.1637	11.2755
USA	1975	4.5597	6.8412	9.8601
USA	1985	5.7762	7.4436	9.6133
USA	2003	5.7494	7.6288	9.8775
USA	2004	6.0747	7.8297	9.9176
USA	2005	5.6025	7.2533	9.1768
USA	2006	5.7324	7.2808	9.1530
USA	2007	5.6225	7.0854	8.9053
USA	2008	5.5204	6.9443	8.7199
USA	2009	5.3823	6.7132	8.4058
USA	2010	5.0263	6.2237	7.8291

These three measures provide very similar results to other studies of the economic value of a life, however, our measures do not include childhood or retirement. This methodology provides a simple approach to estimate the value of the remaining years of one's life or really any function of time that is of interest. While it is not the focus of this

paper, this approach could be used as an alternative in calculating the value of human life in future studies.

The estimate, R , is a measure of discounted lifetime income, which can be a difficult concept to understand. It is a quantity that is not observed, but it is interesting to examine the role of income inequality in determining this measure. In order to do this, we regressed R on the Gini coefficients and its square for each country, which gave the following results.

$$R_{it} = -17.41 + 119.34 \text{ gini}_{it} - 149.86 \text{ gini}_{it}^2 + \varepsilon_{it}$$

(6.41) (35.52) (47.31)

Of course, this is a concave function of the Gini coefficient variables, which are significant at the 95% level; so, it can be maximized with a Gini coefficient of 0.3982 with the maximum being attained at \$6.35 million dollars. The US and Spain are the two countries that came closest to this value in 2003 and 2004, respectively, with both countries showing strong economic growth in those years. This indicates that too much or too little inequality devalues the average value of life. This suggests that income inequality might be a quantity that governments could consider when attempting to regulate the economy. It is quite interesting to note that this value is not significantly different from the square of the inverse golden ratio. Of course, this relation could also be explored in further research.

The table below gives the results for the Marshallian wage elasticity for all of the countries for which we have all of the relevant variables. The fourth column refers to the elasticity in the basic model that is a series of static optimal optimization problems and

the fifth column refers to the basic model in a fully dynamic setting. The fifth column and sixth columns refer to the static and dynamic home production model.

Marshallian Wage Elasticity

Country	Year	Tax Rate	Static	Dynamic	Static HP	Dynamic HP
Canada	1981	0.3429	0.0184	0.1663	0.0679	0.6687
Canada	1986	0.3814	0.098	0.331	0.2707	0.606
Norway	2000	0.5221	0.0211	0.3742	0.4302	0.7578
Germany	1991	0.5425	0.053	0.0949	-0.0485	1.0175
Germany	1992	0.5531	0.072	0.5407	-0.118	0.9664
Germany	2001	0.562	0.0844	0.1636	0.2129	0.911
Germany	2002	0.5643	0.0673	0.1582	0.0096	0.9333
Spain	2002	0.4335	0.082	0.0477	0.4558	0.7627
Spain	2003	0.4353	0.1328	0.1909	0.6454	0.9603
UK	2000	0.463	0.1785	0.1076	0.6496	1.1789
UK	2001	0.463	0.1866	0.1806	0.782	1.2359
USA	1975	0.3119	0.0558	0.0728	0.0716	0.8303
USA	1985	0.3262	0.0456	0.0273	0.2301	0.6383
USA	2003	0.3169	0.056	0.0637	0.3346	0.8907
USA	2004	0.3152	0.1092	0.0822	0.4001	0.8589
USA	2005	0.3266	0.0491	0.2117	0.3668	1.1256
USA	2006	0.3347	0.0143	0.4167	0.3479	0.8585
USA	2007	0.3433	0.0445	0.1657	0.4265	0.9624
USA	2008	0.3433	0.0486	0.097	0.3807	0.9728
USA	2009	0.3433	0.028	0.2933	0.3817	0.7794
USA	2010	0.3433	0.0342	0.2459	0.4014	1.0226

This table shows that all of the Marshallian wage elasticities are positive except for the notable exception of Germany 1991-1992. For that survey age was reported in five-year bands, so our data set was considerably smaller after the data was aggregated by age. The general trend in these results is that the static model will have the smallest elasticity. Once we allow labor to be substituted across time as in the dynamic model, we get an increase in elasticity again. The static home production model allows market

consumption to be substituted for home production, which usually results in an even larger elasticity. Finally the wage elasticity of the dynamic home production model is uniformly larger than all of the other models.

Hicks Wage Elasticity

Country	Year	Tax Rate	Static	Dynamic	Static HP	Dynamic HP
Canada	1981	0.3429	0.0334	0.1677	0.2795	0.6745
Canada	1986	0.3814	0.173	0.3439	0.5027	0.6336
Norway	2000	0.5221	0.0312	0.3757	0.6573	0.7609
Germany	1991	0.5425	0.0708	0.0964	0.1409	1.033
Germany	1992	0.5531	0.0947	0.5485	0.072	0.9802
Germany	2001	0.562	0.1122	0.1641	0.4123	0.9139
Germany	2002	0.5643	0.0869	0.1586	0.1883	0.936
Spain	2002	0.4335	0.1341	0.048	0.6962	0.7654
Spain	2003	0.4353	0.2172	0.192	0.9721	0.9634
UK	2000	0.463	0.2478	0.1081	0.8419	1.1826
UK	2001	0.463	0.2622	0.1814	1.0128	1.2399
USA	1975	0.3119	0.1834	0.0738	0.4395	0.842
USA	1985	0.3262	0.0846	0.0275	0.4587	0.6437
USA	2003	0.3169	0.1059	0.0642	0.6538	0.8978
USA	2004	0.3152	0.1857	0.0828	0.6965	0.8651
USA	2005	0.3266	0.085	0.2131	0.6576	1.1335
USA	2006	0.3347	0.0247	0.4195	0.6033	0.8641
USA	2007	0.3433	0.076	0.1668	0.7117	0.9688
USA	2008	0.3433	0.0832	0.0977	0.6628	0.9792
USA	2009	0.3433	0.0478	0.2952	0.6455	0.7844
USA	2010	0.3433	0.0591	0.2476	0.6677	1.0289

Once the income effects have been removed, the static home production model becomes much closer to the dynamic home production model. Considering that the difference in the way the two elasticities are calculated is highly dependent on the scale of the income measurement. In the graphs that follow, this effect becomes clear. The trajectory of the static home production model appears to move closer to the trajectory of the dynamic home production model. As previously discussed, the elasticity of labor

supply with respect to a change in the tax rate will not be the simple labor wage elasticity, but rather the labor wage elasticity rescaled

$$\frac{\partial h}{\partial \tau} = \left(\frac{-\tau}{1-\tau} \right) \frac{\partial h}{\partial w}$$

The table below shows the Marshallian elasticity with respect to the tax rate. Of course the elasticity with respect to the tax rate will be of opposite sign to the wage elasticity, so we report the absolute value of the elasticities to avoid confusion.

Absolute Value of the Marshallian Tax Elasticity

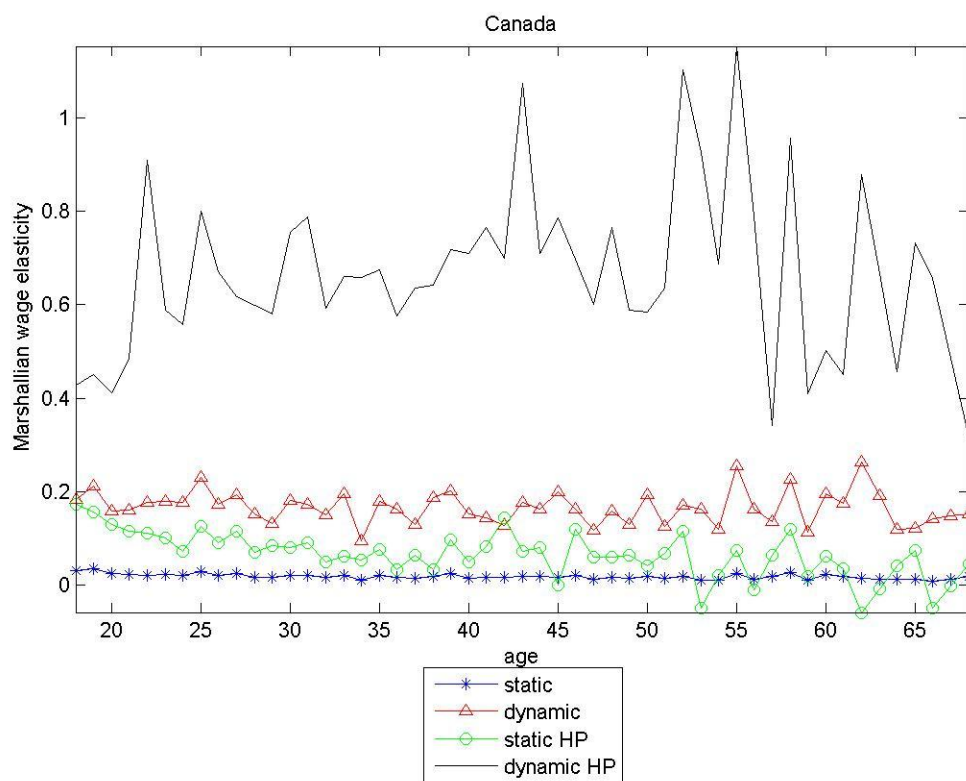
Country	Year	Tax Rate	Static	Dynamic	Static HP	Dynamic HP
Canada	1981	0.3429	0.0096	0.0868	0.0354	0.3489
Canada	1986	0.3814	0.0604	0.204	0.1669	0.3736
Norway	2000	0.5221	0.0231	0.4088	0.4699	0.8278
Germany	1991	0.5425	0.0629	0.1125	0.0575	1.2066
Germany	1992	0.5531	0.0892	0.6692	0.146	1.1962
Germany	2001	0.562	0.1083	0.2098	0.2732	1.1688
Germany	2002	0.5643	0.087	0.2043	0.0125	1.2052
Spain	2002	0.4335	0.0628	0.0365	0.3488	0.5837
Spain	2003	0.4353	0.1024	0.1472	0.4976	0.7404
UK	2000	0.463	0.1539	0.0928	0.5601	1.0164
UK	2001	0.463	0.1609	0.1557	0.6742	1.0655
USA	1975	0.3119	0.0253	0.033	0.0325	0.3764
USA	1985	0.3262	0.0221	0.0132	0.1114	0.309
USA	2003	0.3169	0.026	0.0295	0.1552	0.4132
USA	2004	0.3152	0.0503	0.0379	0.1842	0.3954
USA	2005	0.3266	0.0238	0.1027	0.1779	0.546
USA	2006	0.3347	0.0072	0.2096	0.1751	0.432
USA	2007	0.3433	0.0233	0.0866	0.223	0.5032
USA	2008	0.3433	0.0254	0.0507	0.1991	0.5086
USA	2009	0.3433	0.0146	0.1534	0.1996	0.4075
USA	2010	0.3433	0.0179	0.1286	0.2099	0.5347

The payoff of having low taxes is that it shrinks the labor wage elasticity. Focusing on the dynamic home production model, the Marshallian wage elasticity in the USA was greater than unity in 2010, but the small tax rate results in a labor tax elasticity that is almost half of the wage elasticity. Since Germany has tax rates in excess of 0.5, it must be the case that the tax elasticity is greater than the wage elasticity. In fact the wage elasticity is slightly below unity for Germany, but the labor tax elasticity is inflated to a position greater than unity

Absolute Value of the Hicks Tax Elasticity

Country	Year	Tax Rate	Static	Dynamic	Static HP	Dynamic HP
Canada	1981	0.3429	0.0174	0.0875	0.1458	0.3519
Canada	1986	0.3814	0.1066	0.212	0.3099	0.3906
Norway	2000	0.5221	0.0341	0.4104	0.7179	0.8311
Germany	1991	0.5425	0.084	0.1143	0.1671	1.2249
Germany	1992	0.5531	0.1173	0.679	0.0891	1.2133
Germany	2001	0.562	0.1439	0.2105	0.529	1.1725
Germany	2002	0.5643	0.1122	0.2049	0.2432	1.2087
Spain	2002	0.4335	0.1026	0.0367	0.5328	0.5857
Spain	2003	0.4353	0.1675	0.148	0.7495	0.7428
UK	2000	0.463	0.2137	0.0932	0.7258	1.0196
UK	2001	0.463	0.2261	0.1564	0.8731	1.0689
USA	1975	0.3119	0.0831	0.0335	0.1992	0.3817
USA	1985	0.3262	0.041	0.0133	0.2221	0.3116
USA	2003	0.3169	0.0491	0.0298	0.3033	0.4165
USA	2004	0.3152	0.0855	0.0381	0.3207	0.3983
USA	2005	0.3266	0.0412	0.1034	0.319	0.5499
USA	2006	0.3347	0.0124	0.2111	0.3035	0.4348
USA	2007	0.3433	0.0397	0.0872	0.3721	0.5065
USA	2008	0.3433	0.0435	0.0511	0.3466	0.512
USA	2009	0.3433	0.025	0.1543	0.3375	0.4101
USA	2010	0.3433	0.0309	0.1295	0.3491	0.538

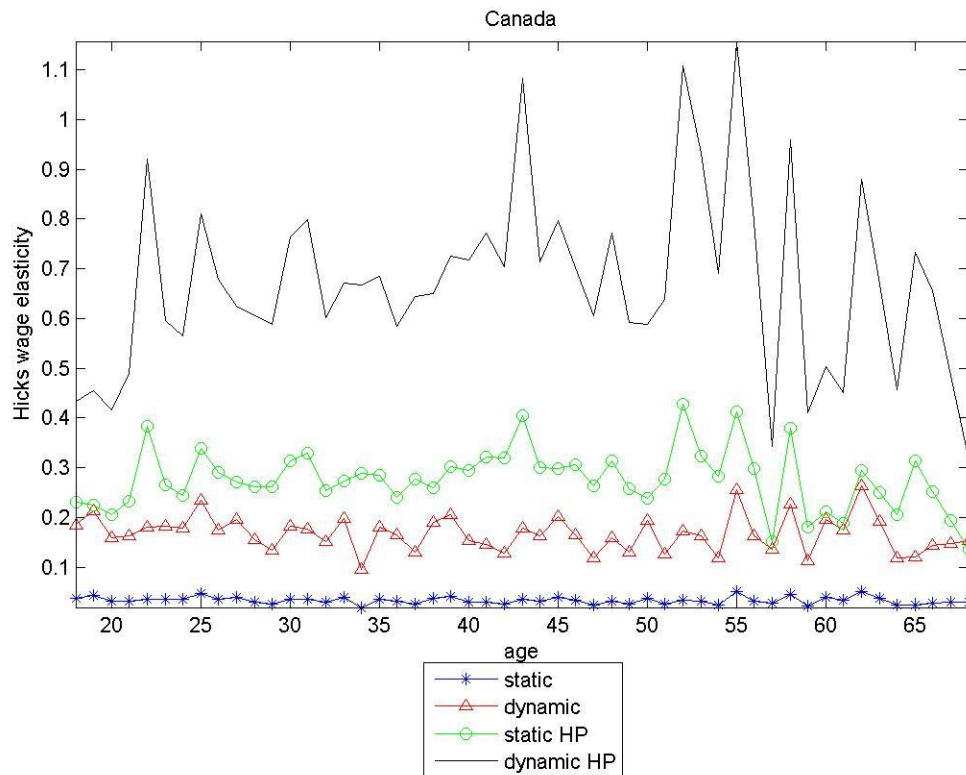
The UK has the largest wage elasticities and the second largest tax burden. As a result the tax elasticity for the UK remains greater than unity and does not shrink very much. For the year 2000, difference between Germany and the UK's tax burden was 0.1, but the factor that is applied to the wage elasticity has a difference of 0.42. This is an example of high taxes creating nonlinear distortions in the labor market. The relationship



between the four models can be seen more clearly through a graphical approach. The relationship between the four models becomes much clear through a graphical approach. In the following, the graphs for the tax elasticities are not included because they are same as the wage elasticity graphs except for rescaling. The data for Canada was only available for the years 1981 and 1986, so below we report the average of the two surveys. The

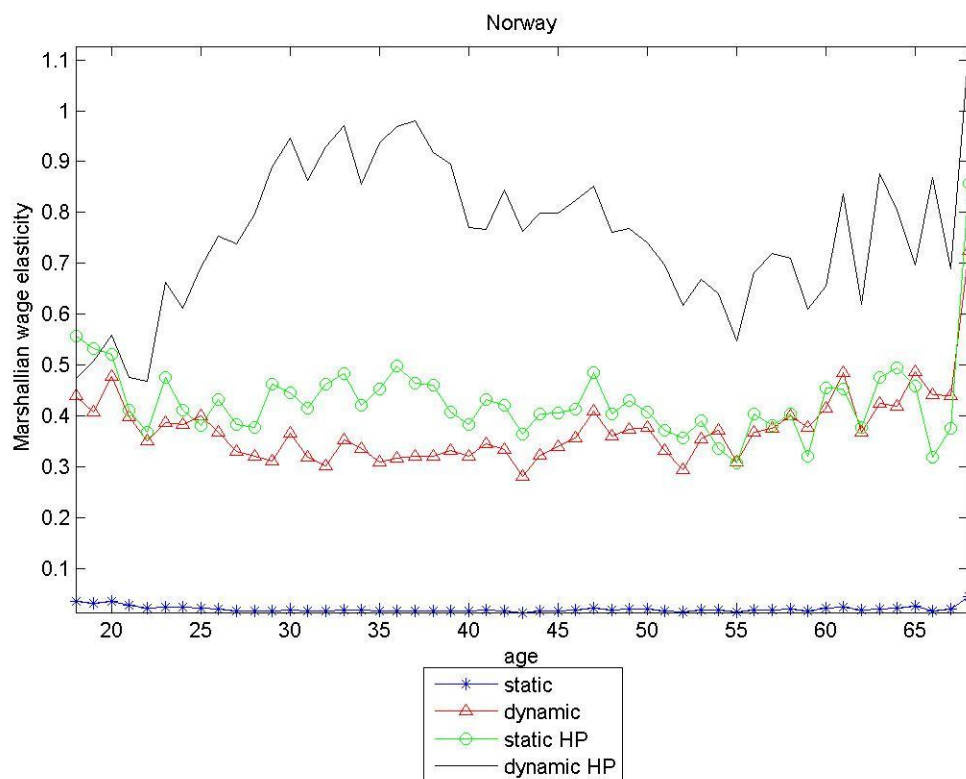
static model doesn't include savings or home production. The dynamic model includes savings, but no home production and so forth.

Here we can see an example of the static home production, HP, model being below the dynamic model and there are even some points in the cohort groups that have the static HP model falling below the static model. There is definitely some similarity in the trajectory of all the models except for the static model, which is essentially constant over the life cycle. It appears as though the trajectory of the dynamic HP is most similar to the dynamic model without home production; however, once the income effects have been removed, as in the graph below, a different pattern emerges.



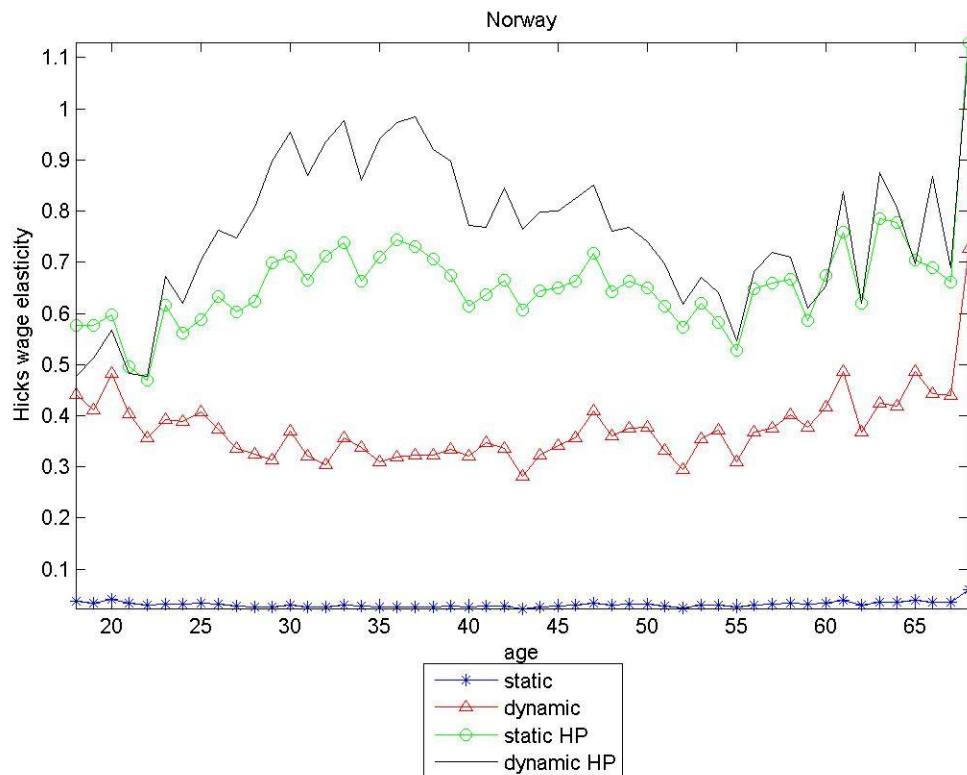
Here we can see that the static HP model moves above the static and dynamic models that don't include home production. The shape of the static HP model also becomes more similar to the dynamic HP model. Overall all four models share similar peaks and valleys with the magnitude and total variation of the elasticity increasing with the complexity of the model.

Canada is a North American country with relatively low taxes, so it is interesting to compare it to a European country with relatively high taxes such as Norway. Norway is also the only European country in this study that is not in the European Union.



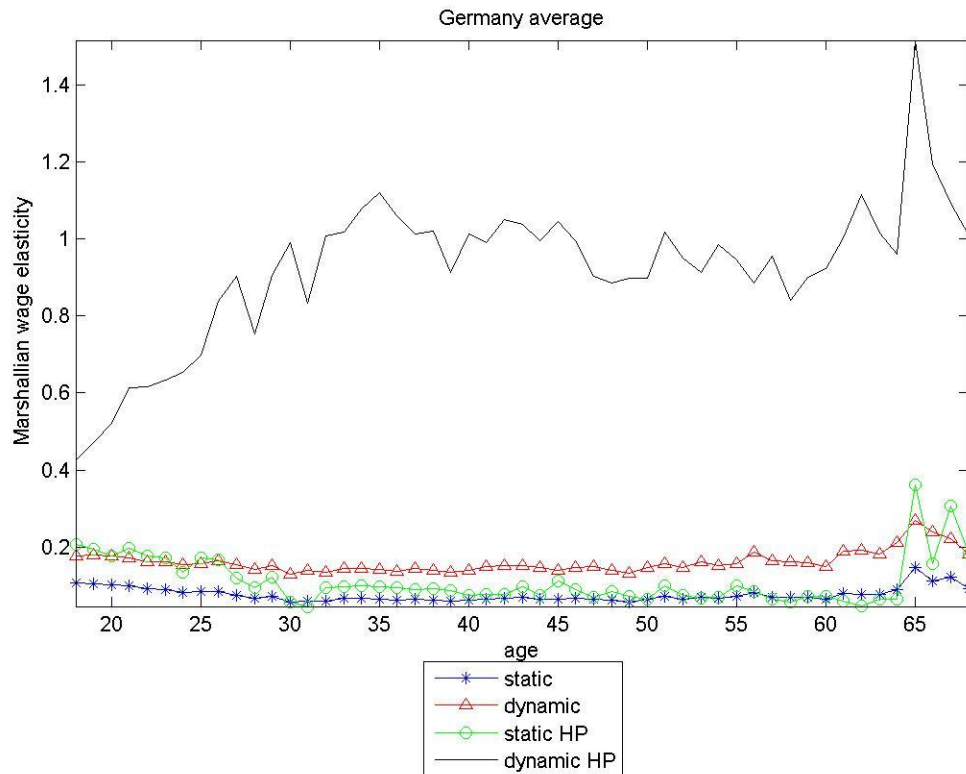
In the graph above, the static HP model is in between the dynamic HP model and the basic dynamic model for the most part. This is slightly different from the Canadian graph where the static HP model was dominated by the basic dynamic model, which is

too say the entire static HP trajectory lies above the basic dynamic model. A property that is pervasive throughout these graphs is that there is a marked increase in wage elasticity as workers approach retirement with the most elastic period of labor supply being in the last period or very near it. In the table below, the static HP model is essentially translated vertically towards the dynamic HP model.



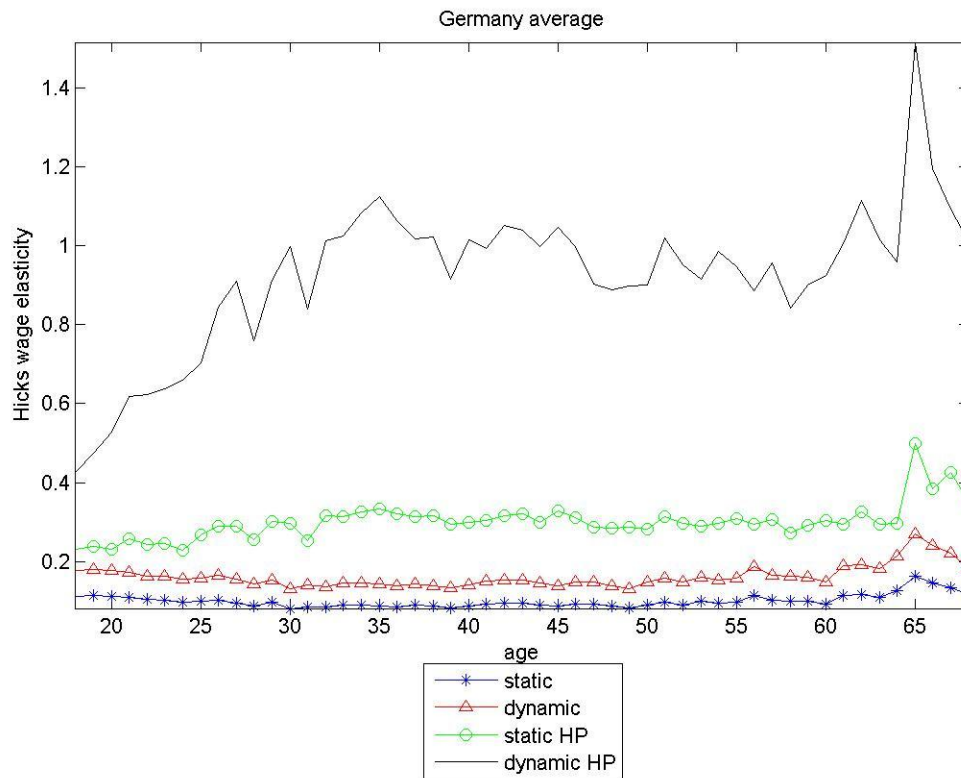
According to the theory of the model, the dynamic HP model should dominate all of the models, but in the first couple years of the working life this relation clearly doesn't hold, however, it is assumed that this is due to noise in the data set. For virtually all of the data points in all of the graphs, the dynamic HP model will dominate the static HP model and the dynamic model will dominate the static model. While the static model appears to be constant around zero, there are some small hills and valleys that can be seen

in the other models. The tax burden in Norway is second only to Germany, which will be examined in the graph of Marshallian wage elasticity presented below.



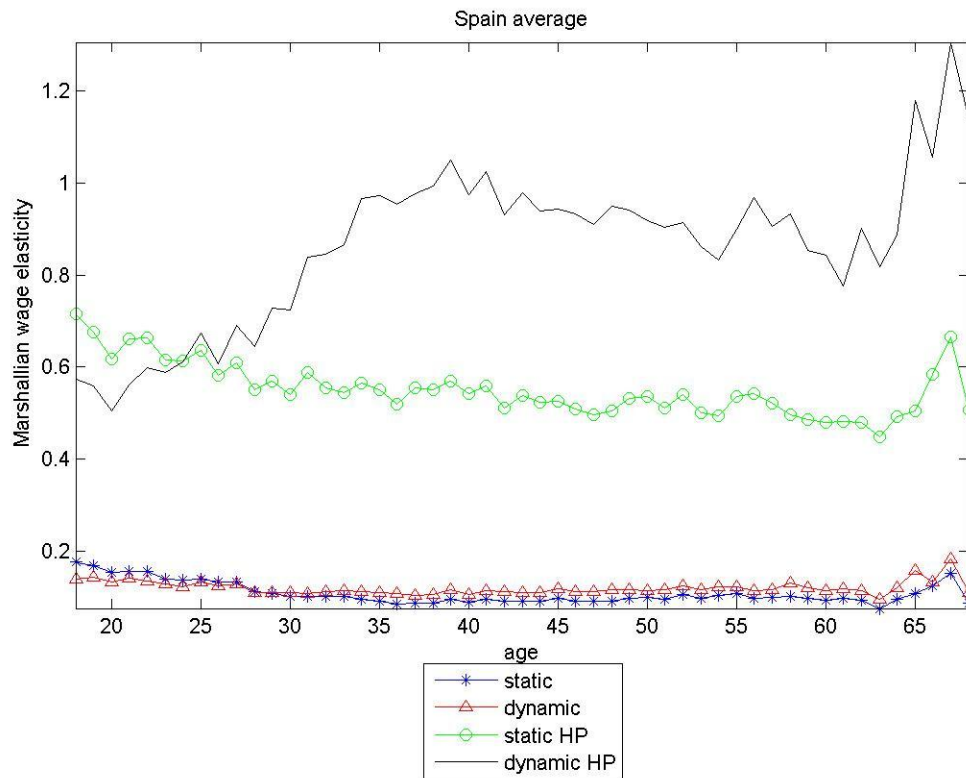
Here is another example of the static HP model being dominated by the dynamic model when the income effects are present. The general trend of the dynamic HP model is increasing then it flattens out with a sharp increase as retirement is approached. The other three models have a very flat appearance with a slight horseshoe trajectory.

There is a vast difference between the dynamic HP model and the other models. The dynamic HP model has the intuitive property that labor supply is inelastic early in the career. The labor supply is roughly unit elastic in the middle part of the career with a transition into elastic labor supply as retirement approaches.



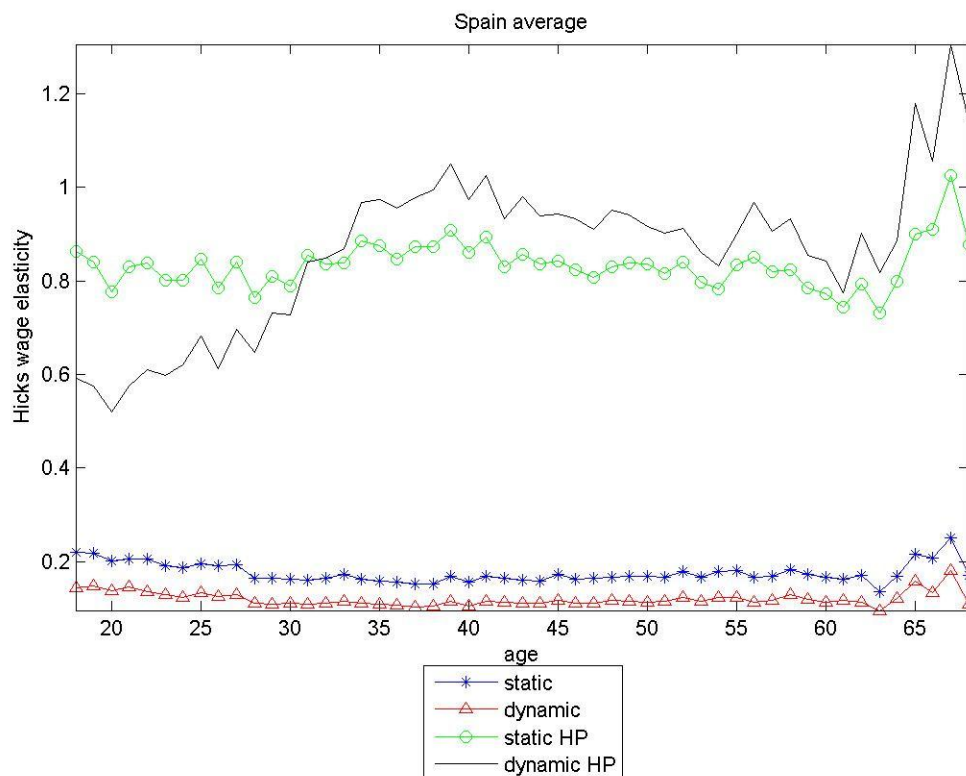
As with some of the previous graphs, removing the income effects results in the static HP model dominating the dynamic model. However, the dynamic HP trajectory still displays the property of inelastic labor supply early career and, roughly, unit elastic labor supply during the middle of the career with a sharp increase near retirement.

While Spain's economy is currently doing quite poorly, the surveys in this study date back to 2002-2003, which is when Spain's economy was thriving well before the housing bubble burst. In the graph below, the dynamic HP model displays a similar trend as in Germany concerning the transition of labor supply from inelastic to quite elastic.



There is an unfortunate crossing between the home production models. It also occurs in the model without home production. This may be due to noise in the data or it may have to do with a small sample in the aggregation of age cohorts because this crossing only occurs during the "college" years.

In the graph below, it can be seen that the crossing still occurs for the home production models, but it is eliminated in the models without home production. This is a bit concerning because the removal of the income effects usually resolves these issues, but this issue seems quite persistent, so this issue might be pursued in future research. During this time period Spain was approaching full employment during the housing boom, which could lead to younger workers to have a more elastic labor supply.

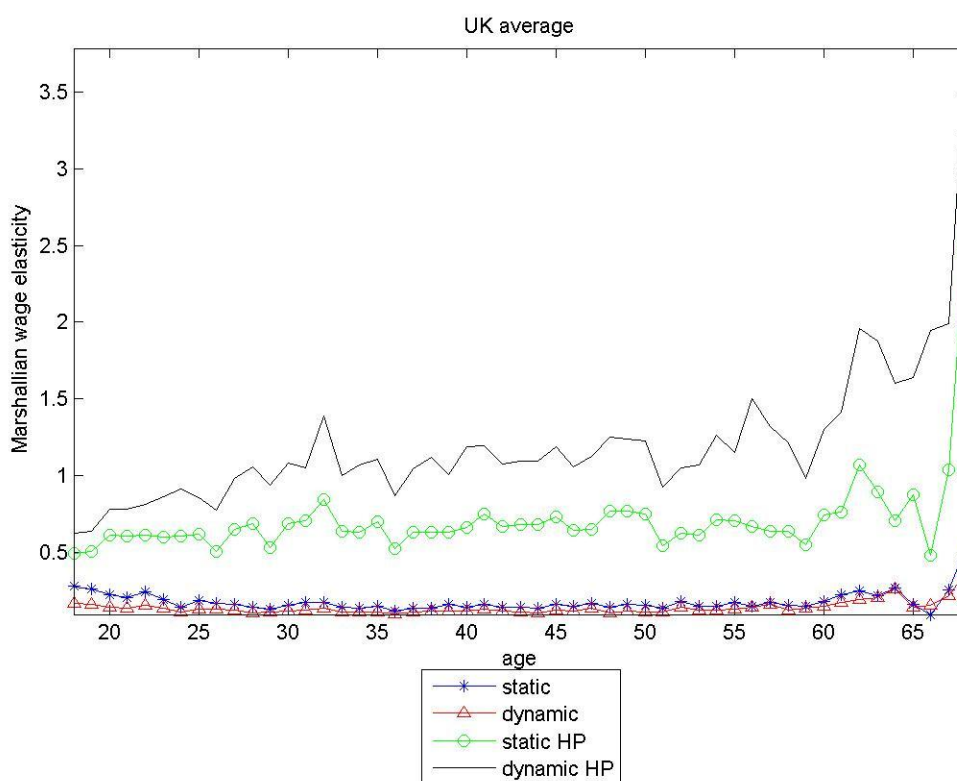


The UK is the last member of the European Union that will be considered; however, the UK does not use the Euro as its currency, so it is interesting to see how the UK labor supply elasticity compares to the other European countries.

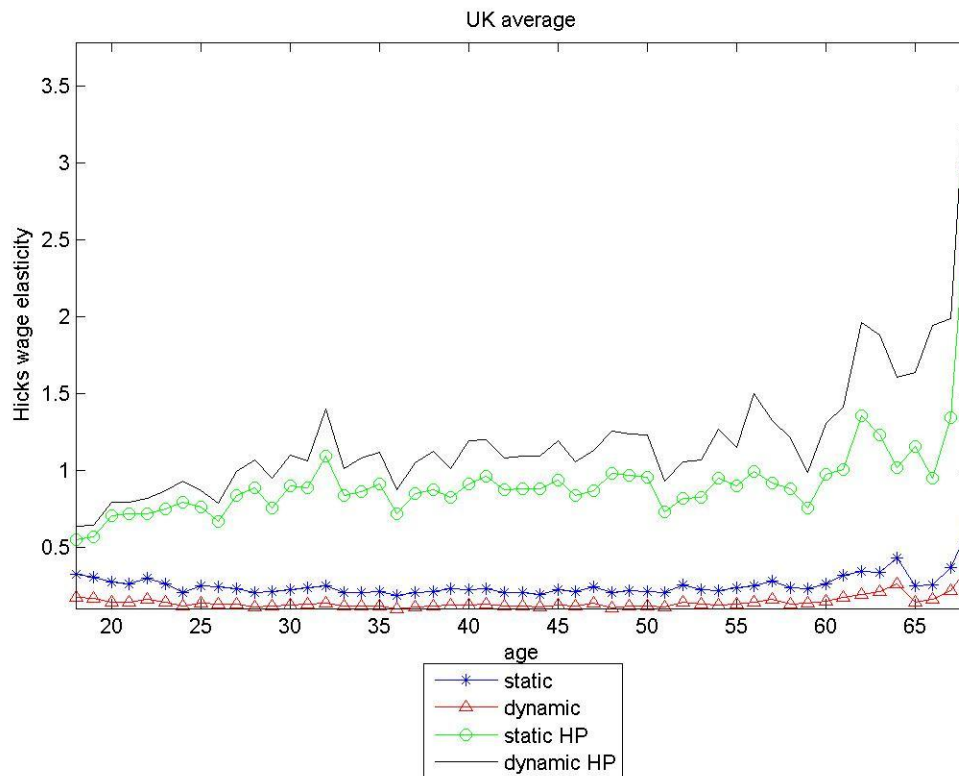
In the Marshallian wage elasticity graph below, there is a very sharp spike in the elasticity as retirement approaches. In fact all of the models show a strong peak in the early sixties followed by a drop and a huge spike after age 65. This is markedly different behavior than with the other countries in the study.

In all of the countries except for the UK, government retirement benefits do not begin until age 65; however, in the UK, women are eligible for retirement benefits at 60. It could be the case that the first spike corresponds to female retirement and the second

spike refers to male retirement. If the two genders were treated equally, we might expect that the 2 spikes average to one smaller spike after age 65.

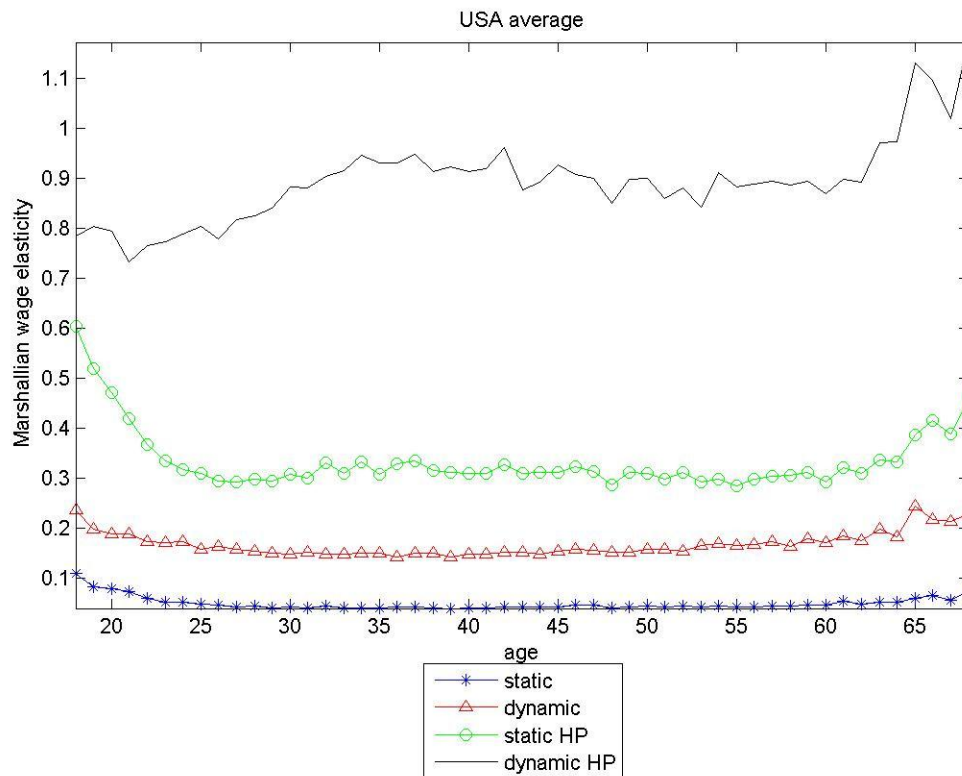


As expected the dynamic HP model and the static HP show generally similar trajectories of differing magnitudes. Unexpectedly the static model dominates the dynamic model. While they are essentially the same, there is still a persistent gap between the two models.



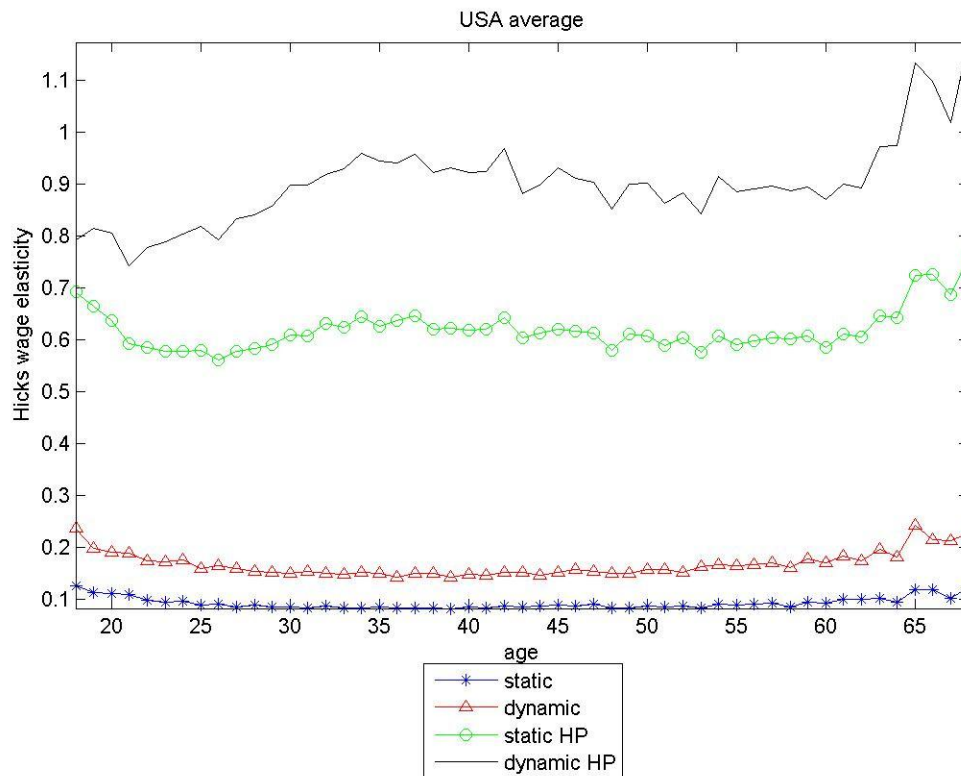
After removing the income effects, the gap between the static and dynamic models has increased. While it may appear that the dynamic model is nonpositive at some points in the trajectory, it is in fact positive throughout the trajectory. One explanation is that home production is a relevant variable, so the parameter estimates for both the static and dynamic models suffer from omitted variable bias, so inference is negatively affected. Meanwhile the gap between the dynamic HP trajectory and the static HP trajectory has decreased as with the other countries.

The US is a particularly interesting example because there is far more data for the US as compared with the other countries in the study.

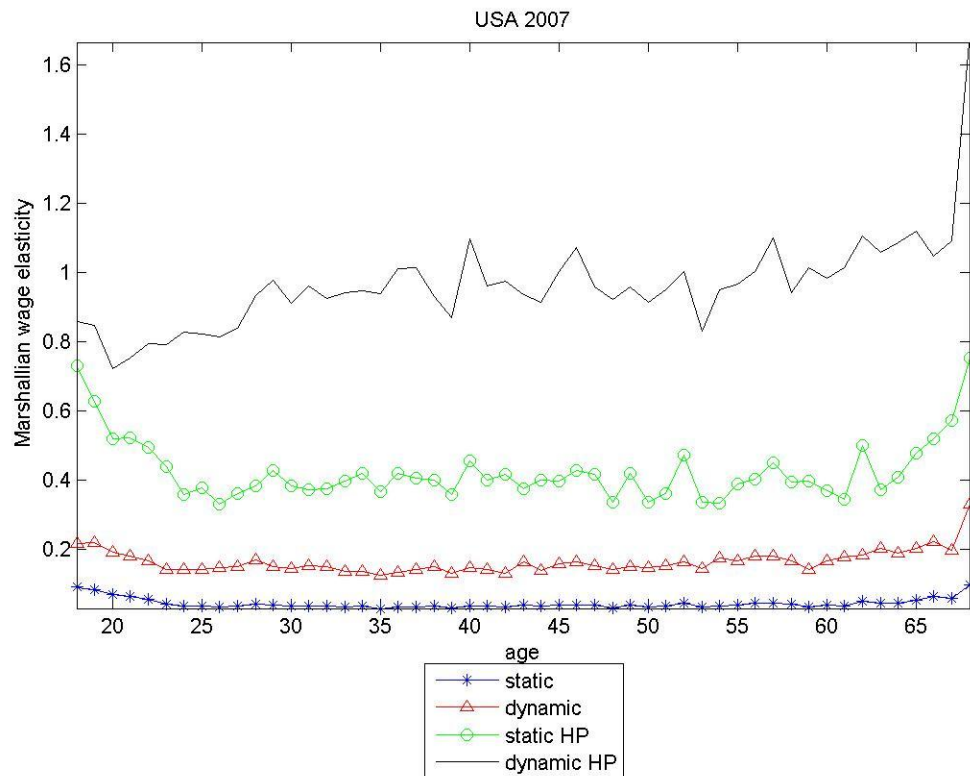


The static HP model and the dynamic model exhibit a horseshoe shape. Meanwhile the dynamic HP is inelastic until retirement the early sixties, but the elasticity does seem to stabilize slightly below one before it jumps into an elastic state towards the end. While the static model is essentially zero throughout the career, it is positive at all ages. There is a small spike in all of the models at age 65, which can be explained by the government benefits that begin to accrue at that age.

The graph below shows similar behavior in the trajectories of the different models. As usual there is large shift in the static HP model towards the dynamic HP model.



There is very little change in the trajectory of the static and dynamic models over the career cycle as compared with the home production models, but all of the models share a similar spike in elasticity at age 65. As discussed previously, this spike is attributed to the government and private benefits that begin to accrue at 65. It is much more intuitive that there should be significant changes in the wage elasticity over time. The difference in wage elasticity for a 65 year old is almost the exact same as an 18 year old worker. It is quite a leap of faith to insist that 18 year olds and 65 year olds respond to a wage change in the same fashion. These graphs have had much less variability in the trajectories because for the US there were many more surveys to construct trajectories, so the elasticity graphs for 2007 are presented below.



The total variation of the trajectories has increased by a significant amount, but the conclusions remain essentially unchanged. The dynamic HP trajectory still displays the property of inelastic labor supply early career and, roughly, unit elastic labor supply during the middle of the career with a sharp increase near retirement. The horseshoe shape remains and is amplified in 2007. It is interesting to note that the spike at age 65 is not present, but it is present in the rest of the US plots. This difference could be attributed to a delay in retirement due to the financial crisis that began in 2007. The dynamic HP model provides results that are a bit larger than standard approaches, but the results are more intuitive and informative.

Conclusions

In this paper, an alternative approach had been applied to dynamic optimization problems. By using the method of repeated substitution, the dynamic optimization problem can essentially be solved as far as the econometrician is concerned. The solution provided will give all of the choice variables in terms of the parameters, labor supply, and the borrowing position. While the dynamics of the borrowing position have been ignored in this paper, it should be solvable in future research.

As discussed earlier unemployment and retirement can be deduced incorporated quite easily to the model. Since a person cannot be unemployed and retired, we can write the expected labor supply in time t with respect to time $t-1$ expectations as

$$E_{t-1}[h_{mit}] = [1 - F_t(\text{unemployment}) - G_t(\text{retirement})]E_{t-1}[h_{mit} | h_{mit} > 0]$$

where F_t and G_t represent the probability of unemployment and retirement, respectively.

The same approach can be taken, however, deducing the wage elasticity will become quite difficult. The benefit of this approach is that it could give insight into the roles inflation, wages, and consumption effects the dynamics of unemployment and retirement.

Even the simple dynamic model produces an alternative method of measuring the value of a workers life. If attention is restricted to the case where only wages and transfers affect the value of life, this is R quantity analyzed previously, similar results are observed with other studies. This measure determines the consumption path for each generation, so increasing R means that the consumption path increases as well corresponding to growth in the economy. This measure is also correlated with the Gini coefficients of the countries under study. The interesting result is that there is an optimal

of the Gini coefficient that maximizes R and that this value, 0.39. The interpretation is that there needs to be a certain amount of income inequality to drive the economy, so there can be such a thing as too much income equality. From time to time communist countries have experimented with excessive income equality resulting in very poor economic results. If leisure and home production is considered to have value to the person, viewed through the lens of opportunity cost, the value of a human life more than doubles. A number of arguments could be made for the inclusion or exclusion of leisure and home production in measuring the value of life.

The static HP model produces labor supply elasticities that are generally larger than the basic static and dynamic models. This is attributed to the increase in substitutes as the model becomes progressively more complicated. The difference in labor elasticities produced by the static HP model and the basic dynamic model is largely due to the ease with which consumption and labor can be substituted with home production or across time. If labor elasticity for the static home production model is greater than the basic dynamic model, then it is cheaper to substitute labor within a time period to home production as opposed to working more in the later stages of the life cycle. If workers are allowed to substitute labor across time and within a time period to home production, as in the dynamic HP model, then a nonlinear jump is observed in the wage elasticity.

The elasticity trajectories provided by the dynamic HP model provide a much more intuitive result. It should seem odd that 18 year olds and 65 year olds respond to wage changes in essentially the same manner. It would seem that young workers are more insensitive to wage changes and have less freedom to change their labor supply,

while older workers can have more freedom to change amount of hours worked and they can change their retirement date depending on market changes.

This is precisely the pattern that emerges in the dynamic HP model. Over the life cycle, labor supply begins in an inelastic state becoming more elastic as experience grows. Then the elasticity roughly constant around one or slightly below with an increase into an elastic state as the retirement age nears.

Returning to taxes, the wage elasticity increases as the tax rate increases, which means the change in the tax rate will be more distortionary in countries that have large tax burdens as opposed to small tax burden countries, even if the change in the tax rate is the same.

However, the wage elasticity is not the relevant notion for a change in the labor supply with respect to a tax change, but it will be the wage elasticity times a nonlinear function of the tax rate. Whenever the total tax burden is greater than 0.5, the tax elasticity will be greater than the wage elasticity. Correspondingly whenever the total tax burden is less than 0.5, the tax elasticity will be shrunk below the wage elasticity. The conclusion being that the total tax burden should be kept below 0.5 to avoid excessive distortions in the labor market.

Finally the dynamic HP model is a more complicated model, but the extra computations result in much more informative results. This approach can be applied to other similar problems and generalized further in future research.

CHAPTER 3: THE NETFLIX COMPETITION

Introduction

The Netflix project was a competition to help solve an information problem. Netflix is a company that rents movies on the Internet. Customers make a list of movies that they are interested in and Netflix mails them those movies as they become available. Then the customer watches the movies and sends it back to Netflix, but Netflix charges only a membership fee without any late fees or rental fees. Thus Netflix can only increase its revenue by getting new members or by getting its old members to upgrade to a more expensive membership. For example, a customer could upgrade from two movies being sent at a time to three movies at a time.

They beauty of being an Internet company is that Netflix has a huge centralized collection of movies that can be distributed cheaply, but that is also the problem: Netflix has so many movies that the customers are flooded with choices. It was easy for customers to search for movie they wanted, but there was no knowledgeable rental store clerk to recommend a good drama like there was in a physical movie rental store.

The solution was to try to make a virtual recommendation system so that people could find movies that were unknown to them or maybe even a forgotten classic. This way Netflix could send the customers recommended movies every time they logged on or added a movie to their rental queue. Hopefully, these recommendations might also be added to the queue and the customer would get more movies they enjoyed. So Netflix allowed its customers to rate any movie in the catalogue, including the movies the customer had rented or browsed. Then the company developed an algorithm called Cinematch that tried to predict which movies the customer might like based on the ratings

of other customers. The idea was to find a movie that John Doe might like, but hadn't already rented by figuring out which other customers were very similar to John Doe. Then for any particular movie that John Doe had never rated or rented, the similar customers could be used as a proxy for John Doe's preferences, thereby allowing Netflix to make recommendations based on the similar customers' preferences.

As Netflix saw it, the quality of its recommendation system was what would make them stand apart from future competitors, so they outsourced it to everyone. They developed the Netflix Prize in which anyone who could beat the Cinematch program by 10% in RMSE, root mean square error, would win a million dollars and publish the results.

It turned out to be quite difficult to reach that 10% improvement and took almost three years. There was no single idea that won the day. The winning team, BellKor, was a blend of algorithms from the most successful teams and there were 107 different estimators used in the winning algorithm. As Abu-Mostafa (2012) points out, even the winning algorithm was only a 10.06% improvement on the CineMatch algorithm. In fact this bound was so tight that the second place team, The Ensemble, submitted a solution that tied the BellKor team, but alas it was submitted 20 minutes too late. While the 10% improvement was chosen rather arbitrarily, it proved to be a monumentally difficult task.

The setup of the contest was to supply the competitors with three things: the quiz set, the probe set, and the training set. The training set contains data on the 480189 customers and 17770 movies. The movie titles are given, but for privacy reasons the customer names are not given. For each customer there is a file that contains all of the ratings that customer has ever made, except for the ones that have been removed for the

quiz set, and the date of that rental. The quiz set is a randomly chosen subset of the training set with the ratings removed. The quiz set is where teams make their predictions and send them to Netflix. Netflix computes the RMSE for the quiz set and sends the results back. Finally, the probe set is another subset of the training set, but this time the ratings are not removed. The idea of the probe set is that teams could practice with a similar dataset in order to hone their algorithms.

There were thousands of teams all over the world using all different types of algorithms. To fix ideas, we present some of the algorithms other teams have used.

As a good first step one might consider using SVD decomposition in order to get a dimension reduction in the problem. Suppose we have an m by n matrix, M , which is very sparse. Then we can use an SVD decomposition to find U and V , so that $M \approx UV'$. The problem is given by

$$\arg \min_{(U,V)} (M - UV')' A (M - UV')$$

where A is a matrix of dummy variables that select only the elements for which we have data (Németh 2007).

This is essentially just a factor model with L factors that estimate our missing data. In general it is unclear how L is to be chosen; however, Kneip, Sickles and Song (2011) present a model that estimates the number of factors by utilizing cubic splines. If we have some panel data set with an endogenous variable Y of size N by T and some exogenous variable X of size P by T , then they consider the model

$$Y_{it} = \beta_0(t) + \sum_{j=1}^p \beta_j X_{ijt} + v_i(t) + \varepsilon_{it}$$

where $\beta_0(t)$ is an average time varying effect and $v_i(t)$ is the time varying effect of individual, i . In order to insure that the model is identified we must also assume that there exists an L -dimensional subspace containing $v_i(t)$ for all $1 \leq i \leq N$ with $\sum_i v_i(t) = 0$. Then they outline a methodology utilizing a spline basis, which is beyond the scope of this paper, to estimate β_1, \dots, β_p and the time varying effects $v_1(t), \dots, v_n(t)$ by minimizing

$$\frac{1}{T} \sum_{i,t} \left(Y_{it} - \bar{Y}_t - \sum_{j=1}^p \beta_j (X_{itj} - \bar{X}_{tj}) - v_i(t) \right)^2 + \frac{\kappa}{T} \sum_i \int_1^T \left(v_i^{(m)}(s) \right)^2 ds$$

over β and all m -times continuously differentiable functions $v_1(t), \dots, v_n(t)$ with $t \in [0, T]$. There is also a test that can be performed on the size of L . This approach is nice because it allows not only the factors to be estimated, but also the number of factors to be used.

While the previous model examined methods to estimate time varying effect, we might also be interested in the spatial relationship of effects. Blazek and Sickles (2010) investigate the knowledge and spatial spillovers in the efficiency of shipbuilding during World War II.

Suppose we are producing a ship, q , through some manufacturing process, which takes L units of labor to produce one ship. In many manufacturing settings, there is an element of learning by doing based on experience, so assume that shipyard, i , in region, j , learns to build ship, h , according to the equation

$$L_{hij} = AE_{hij}^\theta$$

where A is a constant, E_{hij} is the experience of shipyard i in region j that will be used in the production of ship h and $\theta < 0$ represents a parameter ensuring that the number of labor units to produce a single output good is decreasing as experience increases.

Unfortunately experience is not a measurable quantity, so an econometrician might use

the total amount of output as a proxy for experience so that $E_{hij}^O = \sum_{m=1}^{T_h} q_{ijm}$ which is the

total cumulative output of shipyard i up until the time that the production of ship h will begin. However, shipyards within any region are likely to be hiring and firing workers from the same labor pool, which means we should expect experience spillover across shipyards within a region. Of course we can represent this learning spillover within a region as

$$E_{hij}^W = \sum_{m=1}^{T_h} \sum_{n=1}^{I_j} q_{njm} - q_{ijm} = \left(\sum_{n=1}^{I_j} E_{hnj}^O \right) - E_{hij}^O$$

This is just the cumulative experience of the entire region j without the experience of shipyard i so that we capture the effect of the other shipyards in the region. Finally there could also be learning spillovers across regions which can be represented as

$$E_{hij}^A = \left(\sum_{m=1}^{T_h} \sum_{n=1}^J \sum_{o=1}^{I_j} q_{nom} \right) - E_{hij}^W = \left(\sum_{n=1}^J \sum_{o=1}^{I_j} E_{hno}^O \right) - E_{hij}^W$$

In an estimation context, the problem can be represented as a production frontier problem.

$$\ln L_{hij} = \alpha_i + \theta_O \ln E_{hij}^O + \theta_W \ln E_{hij}^W + \theta_A \ln E_{hij}^A + v_{hij}$$

where α_i is the fixed effect of shipyard i and v_{hij} is iid normal with mean zero and variance σ_v^2 . However, this model doesn't incorporate the inevitable inefficiencies that occur in any manufacturing process such as new workers or changes in wages. Blazek and Sickles (2010) model this inefficiency with a nonnegative random variable, μ_{hij} , that represents the organizational forgetting that occurred in the shipyard. Their model is given by

$$\ln L_{hij} = \alpha_i + \theta_O \ln E_{hij}^O + \theta_W \ln E_{hij}^W + \theta_A \ln E_{hij}^A + \mu_{hij} + v_{hij}$$

$$\mu_{hij} = \delta_0 + \delta_1 SR_{hij} + \delta_2 wage_{hij} + \delta_3 HR_{hij} + \varepsilon_{hij}$$

where SR_{hij} is the separation rate of employees during the production of ship h , $wage_{hij}$ is the average hourly wage rate at shipyard i , HR_{hij} is the hiring rate of new workers for ship h in shipyard i and $\varepsilon_{hij} > 0$ is iid truncated normal with mean zero and variance σ^2 so μ_{hij} will be a nonnegative truncation of the normal distribution with variance σ^2 and mean $\delta_0 + \delta_1 SR_{hij} + \delta_2 wage_{hij} + \delta_3 HR_{hij}$. This model seeks to explain the learning that takes place to build ship by comparing firms that close to each other in terms of physical distance. In our model, we will seek to find a measure of distance between customers and movies, but this measure is not a given parameter like distance. This model gives us yet another approach to modeling the interdependent relationships that occur in real world modeling.

One of the most successful approaches to the problem was the neighborhood-based model, (k-NN). Suppose we are trying to predict the rating of movie i by customer u , call it r_{ui} . First we would use some metric, like the correlation between movies, to

choose a subset of the movies, $N(i;u)$, that customer u had already rated that were "close" to the movie in question. For simplicity only the f closest neighbors are kept, we would have the prediction rule

$$r_{ui} = \frac{\sum_{j \in N(i;u)} w_{ij} r_{uj}}{\sum_{j \in N(i;u)} w_{ij}}$$

where w_{ij} represents the similarity between the movie i and movie j and w_i is a vector with f elements. For example, it could just be the correlation between movies. If our similarity measure is 1 whenever the movie is a drama and 0 otherwise, then our estimator will simply be the average of all the drama movies that the customer rated. However, the similarity weights could also be estimated in some fashion.

A more advanced approach tries to estimate the similarity coefficients, which was the BellKor team's approach. The first step in their algorithm was to remove all of the global effects by running the regression

$$Y = X\beta + \varepsilon$$

where Y is a vector of the ratings by users for different movies and X contains global information like movie indicators, time, user indicators, and combinations of the previous. The rest of the analysis will focus on predicting the residual, r_{ui} , from this regression, so our final prediction will be given by

$$\text{prediction}_{ui} = (X\hat{\beta})_{ui} + \hat{r}_{ui}$$

As a way to improve upon the k-NN models, a least squares approach might be taken to minimize the error in our prediction rule. if $U(i)$ is the set of customers that rated movie i , then for each customer $v \in U(i)$ there is a subset $N(i;u,v) \subseteq N(i;u)$ of the

movies that customer v has rated within the neighborhood of customer u . Initially consider the case where all of ratings by person v are known, then the least squares problem can be written as.

$$\min_w \sum_{v \in U(i)} \left(r_{vi} - \frac{\sum_{j \in N(i;u,v)} w_{ij} r_{vj}}{\sum_{j \in N(i;u,v)} w_{ij}} \right)^2$$

This approach gives equal weight to all customers, but we would like to give more weight to customers that are more influential, so they use a weighting function

$c_i = \left(\sum_{j \in N(i;u,v)} w_{ij} \right)^2$ for each user resulting in the following optimization problem.

$$\min_w \sum_{v \in U(i)} \frac{c_i}{\sum_{v \in U(i)} c_i} \left(r_{vi} - \frac{\sum_{j \in N(i;u,v)} w_{ij} r_{vj}}{\sum_{j \in N(i;u,v)} w_{ij}} \right)^2$$

Following Bell (2007), we can rewrite this problem as an equivalent GMM problem subject to nonlinear constraints

$$\min_{w, \lambda \geq 0} w' Q w + \lambda (1 - \sum w_i)^2$$

where $Q_{jk} = \frac{\sum_{v \in U(i)} \delta_{jk} (r_{vj} - r_{vi})(r_{vk} - r_{vi})}{\sum_{v \in U(i)} \delta_{jk}}$ and $\delta_{jk} = \begin{cases} 1 & j, k \in N(i;u,v) \\ 0 & \text{otherwise} \end{cases}$. However, this

approach ignores some information between customers. If we have some measure, s_{jk} ,

of the similarity between customers, then we have the simple modification

$$\delta_{jk} = \begin{cases} s_{jk} & j, k \in N(i;u,v) \\ 0 & \text{otherwise} \end{cases}$$

Previously it had been assumed that the ratings were known, but in reality the number of terms that determine the support for Q_{jk} can vary greatly within the data set, so a shrinkage factor was used

$$\hat{Q}_{jk} = \frac{\sum_{v \in U(i)} \delta_{jk} (r_{vj} - r_{vi})(r_{vk} - r_{vi}) + \frac{\alpha}{f^2} \sum_{jk} Q_{jk}}{\alpha + \sum_{v \in U(i)} \delta_{jk}}$$

where f^2 is the number of elements in Q and α is a shrinkage parameter. Of course we can repeat this process by reversing the roles of customers and movies, but it is less effective, however, the two different results can be combined for further improvements.

Instead of removing the global effects and then estimating the residuals, a refinement can be made that estimates the global effects and the residuals simultaneously. The basic problem is given by

$$\min_{\beta, w, c} \sum_{(v,i)} \left(r_{vi} - \beta_0 - \beta_v - \beta_i - \frac{\sum_{j \in R^k(i;v)} w_{ij} (r_{vj} - \beta_v - \beta_j - \beta_0)}{\sqrt{|R^k(i;v)|}} - \frac{\sum_{j \in N^k(i;v)} c_{ij}}{\sqrt{|N^k(i;v)|}} \right)^2$$

$$\sum_{j=0}^{i+v} \beta_j^2 + \sum_{j \in N^k(i;v)} c_{ij}^2 + \sum_{j \in R^k(i;v)} w_{ij}^2 \geq 0$$

where $R^k(i;v)$ is the set of the k most similar movies to movie v that have available ratings. This set takes into account the information of the levels of the ratings, but information is also available implicitly because the act of rating a movie says provides information that should be utilized. In order to use this implicit information, $N^k(i;v)$ is the set of k most similar movies to movie v that are rated by customer i , even if the actual

rating is unavailable because it is part of the quiz set. Previously the β term was estimated by a fixed effects approach, but it will be driven by the data with this approach.

An alternative to k-NN is a latent factor approach. Paterek, A. (2007) approached the problem by utilizing an augmented SVD factorization model. Under this model, the optimization problem becomes

$$\min_{\beta, p, q} \sum_{(v,i)} (r_{vi} - \beta_0 - \beta_v - \beta_i - q_i^T p_v)^2 + \lambda \left(\|p_v\|^2 + \|q_i\|^2 + \sum_{j=0}^{i+v} \beta_j^2 \right)$$

where this sum is taken over all known customer-movie pairs with known ratings. This turned out to be a very effective approach, but a refinement can be made that includes implicit feedback as in the previous model.

$$\begin{aligned} \min_{\beta, p, q, y} \sum_{(v,i)} (r_{vi} - \beta_0 - \beta_v - \beta_i - q_i^T m_v)^2 \\ m_v = p_v + \frac{\sum_{j \in N^k(i;u)} y_j}{\sqrt{|N^k(i;u)|}} \\ \sum_{j=0}^{i+v} \beta_j^2 \geq 0 \\ \|p_v\|^2 + \|q_i\|^2 + \sum_{j \in N^k(i;u)} y_j^2 \geq 0 \end{aligned}$$

Here implicit information is being applied to the user portion of the matrix factorization, which improves RMSE. To get an idea on how these two different algorithms are combines. This approach still doesn't include the movie-customer interaction term of the SVD model, so the two approaches can be combined into a single optimization problem given below

$$\begin{aligned}
& \min_{\beta, w, c} \sum_{(v,i)} \left(r_{vi} - \beta_0 - \beta_v - \beta_i - q_i^T m_v - \frac{\sum_{j \in R^k(i;v)} w_{ij} (r_{vj} - \beta_v - \beta_j - \beta_0)}{\sqrt{|R^k(i;v)|}} - \frac{\sum_{j \in N^k(i;v)} c_{ij}}{\sqrt{|N^k(i;v)|}} \right)^2 \\
& m_v = p_v + \frac{\sum_{j \in N^k(i;u)} y_j}{\sqrt{|N^k(i;u)|}} \\
& \sum_{j=0}^{i+v} \beta_j^2 \geq 0 \\
& \sum_{j \in N^k(i;v)} c_{ij}^2 + \sum_{j \in R^k(i;v)} w_{ij}^2 \geq 0 \\
& \|p_v\|^2 + \|q_i\|^2 + \sum_{j \in N^k(i;u)} y_j^2 \geq 0
\end{aligned}$$

This is just one example of combining two different algorithms into a single approach. Over the course of the competition, many of the teams collaborated and started to use many different algorithms until the winning algorithm, which was composed of 3 teams and 107 algorithms. One of the most important lessons learned during the competition was the importance of using a diverse set of predictors in order to achieve greater accuracy.

Case based utility is based on the idea that memories of our past decisions and the results of those decisions generate our preferences. That is to say we can represent our preferences as a linear function of our memories. As discussed in chapter 1, Let A be the set of acts that are available to the decision maker from some decision problem p . Also let $c = (a, q, r) \in M$ be the triple consisting of the act, a , chosen in a decision problem, p , and the outcome, r , that resulted from the act. For any given subset of memories, I , preferences can be expressed over acts conditional on those memories, which we denote by $\{\succeq_I\}$. Gilboa and Schmeidler (1995) prove the existence of a utility function given certain regularity conditions which will be represented as

$$U(a|p) = \sum_{(a,q,r) \in M} s(p,q)u(r)$$

So the term $s(p,q)$ is the similarity over the decision problem given that act, a , was chosen. The similarity matrix will not be unique in the sense that the preference structure can be generated by some other similarity matrix, \tilde{s} , that satisfies

$$\tilde{s} = \alpha s + \mu \mathbf{i}'$$

where α is a positive scalar, μ is an arbitrary column vector, and \mathbf{i} is a column vector of ones. If we only consider similarity matrices that have rows summing to unity, then it can be shown that the possible weighting matrices must have the form

$$\tilde{s} = \alpha s + (1 - \alpha) \mathbf{i}(\mathbf{i}' \mathbf{i}) \mathbf{i}'$$

which means $0 \leq \alpha \leq 1$ because all similarity measures must be positive. For our purposes the dimensions of column vectors will be quite large, so all similarity matrices can be approximated by

$$\tilde{s} \approx \alpha s$$

where $0 \leq \alpha \leq 1$. This fact can be used to search for the most accurate similarity in a certain class of similarity matrices.

In order to see the relationship between CBDT and the Netflix problem, suppose r_{vp} is the rating of movie, p , and the rating of this movie is acted out by customer, v , so that in our CBDT language

$$U(v|p) = r_{vp} = \sum_{c \in M} s(p,q)u(r)$$

where $s(p,q)$ represents the similarity between movies p and q . Naturally the result, r , will be the reported rating of movie q acted out by customer v , which means

$$r_{vp} = \sum_q s(p,q)r_{vq}$$

In practice the similarity function will be unknown, any number of similarity functions can be chosen to represent the preference structure. For example, the k most correlated movies could be used as weights in a k -NN type estimate that would give

$$s(p,q) = \frac{\sum_q w_{pq} H(w_{pq} - w_p^{k-1}) r_{vq}}{\sum_q w_{pq} H(w_{pq} - w_p^{k-1})}$$

where w_{pq} is the correlation between movies and w_p^{k-1} is the $k-1$ largest correlation for movie p . This simply means that only the k most highly correlated movies are used to predict the rating. Gilboa and Schmeidler (1995) actually point out that the k -NN approach is a violation of the regularity conditions guaranteeing the CBDT representation of utility. They suggest that all observations be used and simply choose small weights for the less similar cases. In fact this is precisely how the Netflix competitors altered the k -NN approach to produce more precise estimates of customer's movie preferences. Recall the early approach taken by the BellKor team to the Netflix problem.

$$\min_w \sum_{v \in U(i)} \left(r_{vi} - \frac{\sum_{j \in N(i;u,v)} w_{ij} r_{vj}}{\sum_{j \in N(i;u,v)} w_{ij}} \right)^2$$

This can be interpreted as a CBDT optimization where the similarity function is learned from the data. The weights are chosen to minimize MSE and there is no limit on the number of nonzero weights, as there is with a standard k -NN approach.

But we may also have a situation where there is similarity between act decision problem pairs. For example, if our set of acts consists of buying or selling a stock and our set of decision problems consists of buying the stock when the price is high or low, then we may have a situation where "buying the stock when the price is low" is more similar to "selling the stock when the price is high" than "selling the stock when the price is low". Gilboa and Schmeidler (1997) provide axioms that allow a generalization that includes similarity over the pair of decision problems and acts by

$$U(a) = \sum_{(q,b,r) \in M} w((p,a), (q,b)) u(r)$$

This generalization allows for cases and acts to be separated. There are many possibilities, but Gilboa and Schmeidler (1997) provide a multiplicative approach that satisfies the necessary axioms. It was presented as

$$w((p,a), (q,b)) = w_p(p,q) w_a(a,b)$$

Since the weights are positive, the negative logarithm can be taken of both sides to derive an additively separable similarity function given by

$$w((p,a), (q,b)) = w_p(p,q) + w_a(a,b)$$

This is the similarity function that will be used in our model. As before the utility of any result is simply the reported ratings of a movie, so

$$r_{ap} = \sum_{(b,q) \in M} [w_p(p,q) + w_a(a,b)] r_{bq}$$

where r_{ap} is the rating provided by customer, a , for movie, p . Recall that the movie represents the decision problem and the customer represents the act of providing a rating.

This weighting function would have been difficult to implement in practice, so a first order approximation was used.

$$r_{ap} = \sum_q w_p(p, q) r_{aq} + \sum_b w_a(a, b) r_{bp}$$

This weighting function keeps only the most informative movie ratings, which are presumably the ratings made by customer, a , for other similar movies. Similarly the ratings of movie, p , are weighted by the most similar customers. By assumption $w(x, x) = 1$, this fact can be used to rewrite the multiplicatively separable weighting function can be written as

$$r_{ap} = \sum_{(b,q) \in M} w_p(p, q) w_a(a, b) r_{bq} = \sum_{(a,q) \in M} w_p(p, q) r_{aq} + \sum_{(b,p) \in M} w_a(a, b) r_{bp} + \sum_{(b,q) \in M/(a \cup p)} w_p(p, q) w_a(a, b) r_{bq}$$

If the third term is dropped, the equivalence can be seen, but a generalization of our model that will incorporate this functional form will be given later.

$$r_{ap} = \sum_q \lambda_1 w_p(p, q) r_{aq} + \sum_b \lambda_2 w_a(a, b) r_{bp}$$

As discussed previously, we are interested in the class of weighting matrices that have rows summing to unity. As previously demonstrated the weighting matrices will not be unique, but all of the qualifying matrices can be represented with the functional form above as long as the restriction $0 \leq \lambda_1 + \lambda_2 \leq 1$ is imposed. Finally our CBDT based model will have the functional form

$$r_{ap} = \sum_q \lambda w_p(p, q) r_{aq} + \sum_b (1 - \lambda) w_a(a, b) r_{bp}$$

with $0 \leq \lambda \leq 1$. The details of how such a model can be implemented to predict movie ratings in the Netflix competition will be discussed below.

Model

This model began as a class project at Rice University. Naaman, Dingh, and Taylor (2012) used this class project as a springboard to develop a full-scale algorithm for the Netflix competition. Most of the algorithms focused on using only the information between movies because there were so many more customers than users and each movie has much more data than each customer, which is clearly more effective than just using information between customers. We wanted to directly incorporate this symmetry into our algorithm instead of trying to mash two different results together. What set our algorithm apart is that it tries to combine the two sides sort of like digging a tunnel from both sides of the river instead of just one side. However, the trick was to make sure the tunnel met in the middle.

In calculus one can represent a function at any given point by using the first derivative of the function evaluated at some other point suitably close. This concept guided our thinking in that we could take an expansion of the customer's preferences around a particular customer-movie pair by choosing a small neighborhood of customers and movies that were in some sense "close" to that customer-movie pair. This seemed to point us in the direction of spatial regression and case based utility.

Suppose that we have an $N \times 1$ vector of endogenous variables y and that there exists some linear expansion of y_i in terms of the other endogenous variables so that we can write

$$y_i = \sum_j w_{ij} y_j + \varepsilon_i \text{ with } E(\varepsilon_i) = 0$$

In the context of spatial regression, we interpret the weights as being a representation of a point on a map using other landmarks on that map. So it seems reasonable to assume that this is a convex representation, $W \geq 0$ with $\sum_j w_{ij} = 1$ for all i .

These weights might be the result of some other estimation procedure, which is not pursued in this model. Previously examples were given where some sort of weighted least squares subject to constraints estimated the weights. Switching to matrix notation, we can write

$$(I - W)Y = \varepsilon \text{ with } E(\varepsilon) = 0$$

However, this problem is not identified because $I - W$ is not invertible due to the fact that 1 is an eigenvalue of W . But if we assume that there also exists a scaling factor $0 \leq \lambda < 1$, then we are assured that $(I - \lambda W)^{-1}$ exists and we will have the expansion

$$Y = (I - \lambda W)^{-1} \varepsilon = \varepsilon + \lambda W \varepsilon + \lambda^2 W^2 \varepsilon + \dots$$

Under these conditions, we are also assured that $(I - \lambda W)^{-1}$ will have positive entries and the covariance will be given by

$$E(YY') = \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1} \text{ where } E(\varepsilon \varepsilon') = \sigma^2 I$$

This model can easily be extended to the case of an $N \times k$ matrix X of exogenous variables

$$(I - \lambda W)Y = (I - \lambda W)X\beta + \varepsilon$$

The reasoning behind this extension is simply that the exogenous variables should contain the same spatial structure as the endogenous variables in order for the model to make sense. If we assume normality, then $Y - X\beta$ will be multivariate normal with zero

mean and covariance given by $\sigma^2(I - \lambda W)'(I - \lambda W)$. This allows us to write down the log-likelihood as

$$\ln L = -\frac{N}{2} \ln(2\pi\sigma^2) + \ln|I - \lambda W| - \frac{1}{2} \sigma^2 (Y - X\beta)'(I - \lambda W)'(I - \lambda W)(Y - X\beta)$$

Of course, we can solve this minimization problem with standard MLE techniques. However, the $\ln|I - \lambda W|$ term must be computed at each iteration of the nonlinear optimization problem, which can prove to be numerically expensive. Ord (1975) showed that the Jacobian determinant term of the likelihood can be written as

$$\ln|I - \lambda W| = \sum_{i=1}^n \ln(1 - \lambda \rho_i)$$

where $\{\rho_i, 1 \leq i \leq n\}$ is the set of eigenvalues of the spatial weighting matrix. If we appeal to the Schur decomposition of a matrix, in general W will not be symmetric, which allows us to write $W = QSQ^T$ where S is an upper triangular matrix with the eigenvalues of W on the diagonal of S and Q is an orthogonal matrix.

$$\ln|I - \lambda W| = \ln|Q(I - \lambda S)Q^T| = \ln|Q||I - \lambda S||Q^T| = \ln|I - \lambda S|$$

The result follows when we realize that $I - \lambda S$ is an upper triangular matrix with a diagonal entry $S_{ii} = 1 - \lambda \rho_i$, so the result follows and we have much faster computation.

However, Anselin and Hudak (1992) do find evidence for numerical instability for eigenvalues in matrices with more than 1000 entries, but for our purposes all of our weighting matrices will not be so large. This basic model can be extended in many different directions, which are beyond the scope of this paper, but details can be found in Anselin (1988a).

This approach made sense to us because we felt that correlations could be used as a measure of distance between customers and movies. If we had some meaningful weighting matrices that represented the "distance" between movies and customers, then we could apply some spatial regression techniques to reduce RMSE. However, we couldn't find anything in the literature that matched up with our needs, which meant our approach was ad hoc and the result of trial and error.

The first step was to take correlations between movies over the customers that had rated both movies and take correlations between customers over the movies they had in common. However, this does not give a good measure of similarity because two customers may only have one movie in common which tells us little about how similar their preferences are, so the correlations are weighted based on the number of matches two customers or movies had. So if customer u has rated p number of movies and customer v has rated q of the movies that customer u has rated, then the weight for the correlation would simply be $\frac{q}{q+s}$ where s is a scaling factor that could be p or just a parameter. However if the scaling parameter depends on p , then the weighted correlation matrix will be asymmetric because the weight for the correlation between customer u and customer v would still be given by $\frac{q}{q+s}$, but the scaling parameter would depend on how many movies customer v has rated. We decided to leave s as a global scaling parameter so that when z is large the weight will be close to unity and when z is small the weight will scale the correlation down to zero. It made sense that if two customers had rated a large number of movies, but they only had a couple movies in common, then they

were probably not very similar and the high correlations were due to the small sample size.

The next global step of our model was the same as most of the other competitors and that was to get rid of the fixed effects. Let r_{ui} be the rating that customer u gave to movie i with the convention that $r_{u\cdot}$ is the average rating of customer u over all the movies that customer u has rated; and $r_{\cdot i}$ is the average rating of movie i over all the customers that have rated movie i ; and $r_{\cdot\cdot}$ is the grand mean which is the average rating over all customers and all movies. This leaves us with the residuals given by

$$y_{ui} = r_{ui} - r_{u\cdot} - r_{\cdot i} + r_{\cdot\cdot}$$

We also tried a random effects model, but that did not perform as well as the fixed effects approach in terms of out of sample RMSE. Some other panel techniques were also applied, but they provided no improvement in out of sample RMSE. But the rest of our model will be working with the residuals of the fixed effects model.

If we are trying to predict the rating of customer u for movie i , then the next step is to choose a cluster of similar movies and customers. For customer u we choose the c customers that had the largest positive correlation and had rated movie i , but c was capped at thirty in order to make the problem numerically feasible. Repeat this same process for the movie cluster so that we have weighting matrices given by W_c and W_m , which are square matrices of possibly different sizes. The rows of the matrices are ordered by the level of correlation so that the first row of W_c corresponds to customer u ; and the second row corresponds to the customer that is most correlated with customer u ; and the third row corresponds to the customer that is the second most correlated with

customer u and so on. The rows of W_m are ordered in a similar fashion. Finally we standardized the rows of W_m and W_c to sum to unity which is the standard approach in a spatial regression.

We have a panel of $c * m$ data of residuals, which can be stacked giving

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{c1} \\ \vdots \\ y_{cm} \end{pmatrix}$$

If we let $W = \lambda W_c \otimes I_m + (1 - \lambda) I_c \otimes W_m$, then the model can be written

$$Y = WY + \varepsilon$$

In a spatial regression, the parameter λ is a scaling parameter that is restricted to be between zero and one. In that spirit, we also restricted the parameters to sum to less than one, which allows the above representation. If such a representation exists, then the scaling parameter, λ , can be interpreted as the probability the given customer-movie pair can be represented as a weighted sum of the most similar users and $1 - \lambda$ is the probability that the customer-movie pair is represented as a weighted sum of the most similar movies.

Finally we are ready to minimize the log likelihood,

$$L = \ln |I_{cm} - W| - \frac{1}{2} Y' (I_{cm} - W)' (I_{cm} - W) Y$$

where $W = \lambda W_c \otimes I_m + (1 - \lambda) I_c \otimes W_m$. In order to improve the speed of this nonlinear optimization problem, we can again appeal to the Schur decomposition of a matrix. Let $W_c = PT_cP'$ and $W_m = QT_mQ'$ where P and Q are orthogonal matrices with T_c and T_m being upper triangular matrices with the eigenvalues of W_c and W_m on the diagonals, respectively. This allows the $\ln|I_{cm} - W|$ term on the optimization problem to be rewritten as

$$\begin{aligned} \ln|I_{cm} - W| &= \ln|(P \otimes Q)(I_c \otimes I_m - \lambda T_c \otimes I_m - I_c \otimes (1 - \lambda)T_m)(P' \otimes Q')| \\ &= \ln|(I_c \otimes I_m - \lambda T_c \otimes I_m - I_c \otimes (1 - \lambda)T_m)| = \sum_{i=1}^c \sum_{j=1}^m \ln(1 - \lambda\alpha_i - (1 - \lambda)\beta_j) \end{aligned}$$

where $\{\alpha_i, 1 \leq i \leq c\}$ is the set of eigenvalues for W_c and $\{\beta_j, 1 \leq j \leq m\}$ is the set of eigenvalues for W_m and the last equality follows from the properties of upper triangular matrices. This allows for a much quicker numerical computation of the optimization problem.

Once we have our estimator, $\hat{\lambda}$ in hand, then our prediction for the rating that customer u will give to movie i is given by $r_{u..} + r_{.i} - r_{..} + (\hat{W}Y)_1$ and we repeat this process for the next customer-movie pair. We can extend this model by using the rest of the $c * m - 1$ predictions to impute the missing data. When we move on to the next customer-movie pair, there might be some overlap in the predicted values of missing data, but we can simply average the different predictions of the same missing data point. Then we can repeat the whole process using our averaged predictions for the missing data.

Estimation

The algorithm that we used in the Netflix competition begins with forming the fixed effects predictor and residual given by

$$\begin{aligned}\hat{r}_{ui} &= r_{u.} + r_{.i} - r_{..} \\ y_{ui} &= r_{ui} - \hat{r}_{ui}\end{aligned}$$

In general any prediction rule can be used, but the main focus of this approach is to estimate the residual of the predictor.

The next step in the preprocessing is done by finding the weighting matrix, W_m , for the movie effect. Our weighting matrix is computed by shrinking the Pearson correlation, which means an element of the movie weighting matrix will be given by

$$(W_m)_{ij} = \left(\frac{2q_{ij}}{q_{ij} + 480189} \right) \frac{\sum_k (r_{ki} - \bar{r}_{.i})(r_{kj} - \bar{r}_{.j})}{\sqrt{\sum_k (r_{ki} - \bar{r}_{.i})^2 \sum_k (r_{kj} - \bar{r}_{.j})^2}}$$

where q_{ij} is the number of common customers for movie i and j . On the right hand side, we have the Pearson correlation over all of the common customers for movie i and j . On the left, we have our shrinkage factor, which will be 1 whenever $q_{ij} = 480189$.

This means there will be very little shrinkage for a large number of common customers. Conversely the weight will be essentially 0 whenever there are few common customers to support the correlation as a measure of similarity. In a similar fashion, the customer weighting matrix is calculated as

$$(W_c)_{ij} = \left(\frac{2p_{ij}}{p_{ij} + 17770} \right) \frac{\sum_k (r_{ik} - \bar{r}_{i.})(r_{jk} - \bar{r}_{j.})}{\sqrt{\sum_k (r_{ik} - \bar{r}_{i.})^2 \sum_k (r_{jk} - \bar{r}_{j.})^2}}$$

Once these matrices are formed, the last step of the preprocessing is to find the eigenvalues of the weighting matrices so that the log-likelihood can be written as

$$\sum_{i=1}^c \sum_{j=1}^m \ln(1 - \lambda \alpha_i - (1 - \lambda) \beta_j) - \frac{1}{2} Y' (I_{cm} - W) (I_{cm} - W) Y$$

From this point λ can be estimated by a simple line search. Since we are only interested in prediction, the asymptotic distribution of our estimator was not worked out, but bootstrapping the standard errors is feasible, but it was not computationally feasible for our large data set. Once the estimate is in hand, the prediction for the quiz set can be made for the rating of customer c for movie m .

$$r_{cm}^{pred} = r_{c.} + r_{.m} - r_{..} + (\hat{W}Y)_1$$

There will now be $cm-1$ other predictions which can be used to impute the missing data, so this predicted rating will replace any missing data. If there is already a prediction from some other regression then the predictions will be averaged. The number of predictions that have been made will be saved so that the averaging can always be done over all of the predictions that have been made for any missing data point. Of course if the data is not missing, then there is no need to predict it. This whole procedure is then repeated creating a new predicted rating until we achieve convergence in RMSE over the probe set. After convergence is reached, all of the predictions can be submitted for the quiz set.

Results

In order to get an idea about the relationship between the movie and user effects, we iteratively estimated the RMSE of the different fixed effects. First we calculated the RMSE just using the overall mean as our predictor; then we added the movie effect to our estimator; and finally, we estimated the RMSE over the full fixed effects model consisting of customer and movie effects.

Predictor	RMSE	Improvement
Overall mean	1.130	NA
Movie mean and overall mean	1.053	0.077
Customer mean, movie mean, and overall mean	0.984	0.069

The main thing to notice about the table above is that adding the movie effect produces a larger improvement than adding the customer effect to our predictor. Since there are many more customers than movies, it stands to reason that the movie mean is a more robust predictor than the user mean because on average there are many more customer ratings for any given movie than movie ratings for any given user due to the discrepancy between the number of movies and customers.

We wanted to know the effectiveness of the different parts of our model running a linear regression over our actual RMSE that came from running our model over the entire probe.

$$RMSE_j = \sum_{i=1}^9 \beta_i z_{ij} + \varepsilon_j$$

Independent variable	Estimate
Grand mean	– .00010 (0.00100)
Movie mean	– .11608 (0.00102)
User mean	– 0.04138 (0.00107)
Lambda	- 0.02369 (0.00113)
Number of ratings in the $c * m$ data set	– 0.08695 (0.00118)
Average of the first row of W_m	– 0.13912 (0.00194)
Average of the first row of W_c	- 0.03404 (0.00154)
Average of all of W_m except for the first row	0.10803 (0.00193)
Average of all of W_c except for the first row	0.03275 (0.00147)

The user mean was not nearly as effective as the movie mean in explaining a small decrease in RMSE of the model and the grand mean was less effective than both the movie mean and user mean. This falls in line with the experiences of the other

competitors. In fact, most of the competitors used a movie centric approach which exploited the robustness of the between movie effects. The next thing to notice is that the spatial weight, λ , is negatively correlated with RMSE, so there is strong evidence supporting a spatial approach. As expected the number of ratings in the data set is also negatively correlated with RMSE, which simply indicates that the model improves as our data set fills up with actual ratings.

In our model, the single most effective factor was the average of the first row of W_m which is the average weight given to the most similar movies that have been rated by the customer in question. As we have more highly correlated movies, the average of the first row of W_m increases which leads to a decrease in RMSE in terms of conditional expectation. As we saw with the customer mean and movie mean, the movie effect is more relevant than the customer effect, but the model is improved by combining both effects.

It seems counterintuitive that the average entry of W_c and W_m , excluding the first row of both matrices, is positively correlated with RMSE. However, this is really a measure of the similarity between the movies and customers that we are using as a basis to make out estimate for any given customer-movie pair. In essence, there is some overlap in our neighborhood of customers and movies. Ideally we would have a set of movies or customers that is very similar to the rating we are trying to predict, but the set of movies or customers themselves are not very similar to each other.

The final RMSE for our model over the probe set is given in the table below with some naïve models as well.

Predictor	RMSE	Improvement on Cinematch
Overall mean	1.130	-17%
Movie mean and overall mean	1.053	-15%
Customer mean, movie mean, and overall mean	0.984	-2%
Cinematch	0.965	0%
Spatial Model (Probe Set)	0.93	3.76%
Spatial Model (Quiz Set)	0.88	8.8%
Winning Algorithm	0.865	10.006%

This is very close to the 10% improvement that we would need to win the Netflix competition; however, this result is for the probe set which was for practice. When the algorithm was applied to the Quiz set, the RMSE was 0.92 which was a 4.7% improvement on the Cinematch algorithm which was well short of the 10% improvement needed to win. Despite our best efforts, we were never able to explain the discrepancy in accuracy between the two data sets.

Conclusions

In the Netflix competition, we found significant improvements in RMSE by using a spatial model approach that incorporates the interrelationship between movies and customers. One drawback of this approach is its computational difficulty, but the computational burden can still be handled by most data sets that are encountered in the field and by a judicious choice of the maximum size of the neighborhood used to make predictions.

There are also future extensions to this model by using more advanced weighting matrices. One possible future improvement of this model would be to use a more complicated weighting matrix. Recall that the multiplicatively separable model can be written in SBDT form as

$$r_{ap} = \sum_{(a,q) \in M} w_p(p,q) r_{aq} + \sum_{(b,p) \in M} w_a(a,b) r_{bp} + \sum_{(b,q) \in M/(a \cup p)} w_p(p,q) w_a(a,b) r_{bq}$$

In this paper, the final term was dropped resulting in an additively separable model; however, the model can be generalized to a multiplicatively separable model in the following way.

$$W = \lambda W_c \otimes I_m + (1 - \lambda) I_c \otimes W_m - \lambda(1 - \lambda) W_c \otimes W_m$$

$$\ln|I_{cm} - W| = \sum_{i=1}^c \sum_{j=1}^m (1 - \lambda \alpha_i - (1 - \lambda) \beta_j + \lambda(1 - \lambda) \alpha_i \beta_j)$$

where $\{\alpha_i, 1 \leq i \leq c\}$ is the set of eigenvalues for W_c and $\{\beta_j, 1 \leq j \leq m\}$ is the set of eigenvalues for W_m . This approach resulted in a smaller RMSE on a subset of the probe, but it was not pursued any further and may be revisited in future work.

Further improvements seem quite likely if the weighting matrices are estimated via the more complicated methods presented earlier. One could also use a kernel

approach to estimate the similarity matrices, but the asymptotic theory of such an approach are beyond the scope of this paper. This approach can be applied to other panel data sets as another approach to take into account the relationship between effects and aid in making predictions. While this approach may not have been as successful as some of the other Netflix algorithm, we believe that it can provide a refinement to existing recommendation system algorithms. Spatial regression approaches have typically focused on a single dimension of spatial correlation. Our approach has demonstrated a novel approach to estimating models with more than one type of spatial correlation. Finally one of the greatest lessons of the Netflix competition is the emphasis on multiple approaches being combined to provide more accurate predictions of customer's preferences.

REFERENCES

1. Abu-Mostafa, Yaser S. (2012) Machines that Think for Themselves. *Scientific American* 307, pp 78-81.
2. Aguiar, M. and Eric Hurst (2007a). Life Cycle Prices and Consumption. *American Economic Review* 97, pp. 1533-1559.
3. Aguiar, M. and Eric Hurst (2007b). Measuring Leisure: The Allocation of Time Over Five Decades. *Quarterly Journal of Economics* 122, pp. 969-1006.
4. Altonji, J. (1986). Intertemporal Substitution in Labor Supply: Evidence from Micro Data. *Journal of Political Economy*, 94 (June, Part 2), S176-S215.
5. Altug, S. and Robert Miller (1990). Household Choices in Equilibrium. *Econometrica* 58: 543-570.
6. Altug, S. and Robert Miller (1998). The effect of work experience on female wages and labor supply. *Review of Economic Studies*, 65: 45-85.
7. Albanesi, S. and C. Olivetti (2009). Home Production, Market Production and the Gender Wage Gap: Incentives and Expectations. *Review of Economic Dynamics* 12, pp. 80-107.
8. Anselin, L. (1988a). *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.
9. Angrist, J. (1991). Grouped Data Estimation and Testing in Simple Labor Supply Models. *Journal of Econometrics* 47: 243-265.
10. Anselin, L. (1988b). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis* 20, 1-17.

11. Anselin, L. (1992). Space and applied econometrics. Special Issue, *Regional Science and Urban Economics* 22.
12. Anselin, L. and A. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah and D. E. A. Giles, Eds., *Handbook of Applied Economic Statistics*, pp. 237–289. New York: Marcel Dekker.
13. Anselin, L. and S. Hudak (1992). Spatial econometrics in practice, a review of software options *Regional Science and Urban Economics* 22, 509–536.
14. Aten, B. (1996). Evidence of spatial autocorrelation in international prices. *Review of Income and Wealth* 42, 149–63.
15. Baltagi, B. and Dong Li (1999). Prediction in the panel data model with spatial correlation. In L. Anselin and R. Florax (Eds.), *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.
16. Becker, Gary (1965). A Theory of the Allocation of Time. *The Economic Journal* 75, 493-517.
17. Bell, R., and Y. Koren (2007). Scalable Collaborative Filtering with Jointly Derived Neighborhood Interpolation Weights. *IEEE International Conference on Data Mining (ICDM'07)*, IEEE.
18. Bell, R., and Y. Koren (2007). Improved Neighborhood-based Collaborative filtering. *KDD-Cup and Workshop*, ACM press.
19. Bell, R., and Y. Koren, and C. Volinsky (2007). The BellKor solution to the Netflix Prize. http://www.netflixprize.com/assets/ProgressPrize2007_KorBell.pdf.

20. Benhabib, J., Richard Rogerson, and Randall Wright (1991). Homework in Macroeconomics: Household Production and Aggregate Fluctuations. *Journal of Political Economy* 99:1166–87.
21. Blomquist, S. and Urban Hasson-Busewitz (1990). The Effect of Taxes on Male and Female Labor Supply in Sweden. *Journal of Human Resources* 25(3): 317-57.
22. Blundell, R.W., C. Meghir and P. Neves (1993). Labor Supply: an Intertemporal Substitution. *Journal of Econometrics* 59: 137-160.
23. Blundell, R.W., A. Duncan and C. Meghir (1998). Unemployment, Discouraged Workers and Female Labour Supply. *Research in Economics* 52: 103-131.
24. Blundell, R.W. and Thomas H. MaCurdy (1999), “Labor Supply: A Review of Alternative Approaches”, in O. Ashenfelter and D. Cards (eds.), *Handbook of Labor Economics*, Vol. 3, North-Holland, Amsterdam, pp. 1559-1695.
25. Burda, M., Dan Hamermesh, and Philippe Weil (2008). The Distribution of Total Work in the EU and US. In: *Are Europeans Lazy? or Americans Crazy?*, edited by Tito Boeri, Oxford University Press, Oxford.
26. Davis, S. and Magnus Henrekson (2004). Tax effects on Work Activity, Industry Mix and Shadow Economy Size Evidence from Rich Country Comparisons. *National Bureau of Economic Research Working Paper #1059*.
27. Francis N. and Valerie Ramsey (2007). A Century of Work and Leisure. *manuscript, University of California at San Diego*.

28. Freeman, R., and Ronald Schettkat (2001). Marketization of Production and the US-Europe Employment Gap. *Oxford Bulletin of Economics and Statistics* 63, pp. 647-670.
29. Freeman, R., and Ronald Schettkat (2005). Marketization of Household Production and the EU-US Gap in Work. *Economic Policy* 20, pp. 6-50.
30. Gilboa I. and D. Schmeidler (1995). Case-based decision theory, *Quart. J. Econ.* **110**, 605–639.
31. Gilboa, I. and D. Schmeidler (1997). Act-similarity in case-based decision theory, *Econ. Theory* **9**, 47–61.
32. Gilboa, I. and D. Schmeidler (2001). *A Theory of Case-Based Decision*, Cambridge Univ. Press, Cambridge, UK.
33. Gronau, R. (1965). Leisure, Home Production and Work—The Theory of the Allocation of Time Revisited. *Journal of Political Economy* 75, pp. 1099-1124.
34. Hausman, J.J. and P. Ruud (1986). Family Labor Supply with Taxes. *American Economic review (Papers and Proceedings)* 74: 242-248.
35. Heckman, J.J. and T.E. MaCurdy (1980). A Life-Cycle Model of Female Labour Supply. *Review of Economic Studies*, 47: 47-74.
36. Heckman, J.J. (1993). What has been Learned about Labor Supply in the Past Twenty Years. *American Economic review (Papers and Proceedings)* 83: 116-121.
37. Kneip, A., R.C. Sickles, and W. Song (2011). A New Panel Data Treatment for Heterogeneity in Time Trends. *Econometric Theory*.

38. Konstan, J., G. Karypis, J. Riedl, and B. Sarwar (2001). Item-based Collaborative Filtering Recommendation Algorithms. *Proc. 10th International Conference on the World Wide Web*, pp.285-295.
39. Konstan, J., G. Karypis, J. Riedl, and B. Sarwar (2000). Application of Dimensionality Reduction in Recommender System – A Case Study. *WEBKDD'2000*.
40. Koren, Y. (2008). Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model. *Proc. 14th ACM Int. Conference on Knowledge Discovery and Data Mining (KDD'08)*, ACM press.
41. Koren, Y. Factor in the Neighbors: Scalable and Accurate Collaborative Filtering. <http://public.research.att.com/~volinsky/netflix/factorizedNeighborhood.pdf>, submitted.
42. Linden, G., B. Smith and J. York (2003). Amazon.com Recommendations: Item-to-item Collaborative Filtering, *IEEE Internet Computing* **7**, 76–80.
43. Leroux, J., J.A. Rizzo, and R.C. Sickles (2010). The Role of Self-Reporting Bias in Health, Mental Health and Labor Force Participation: a Descriptive Analysis. *Journal of Empirical Economics*.
44. MaCurdy, Thomas (1981). An Empirical Model of Labor Supply in a Life-Cycle Setting. *Journal of Political Economy*, 89, 1059-1085.
45. MaCurdy, Thomas (1983). A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty. *International Economic Review*, 24:2, 265-289.

46. MaCurdy, T., David Green, and Harry Paarsch (1990). Assessing Empirical Approaches for Analyzing Taxes and Labor Supply. *Journal of Human Resources* 25(3): 415-90.
47. McDaniel, C. (2006). Average Tax Rates on Consumption, Investment, Labor and Capital in the OECD: 1950-2003. *manuscript, Arizona State University*.
48. McDaniel, C. (2007). Hours of Work in the OECD 1960-2003: Driving Forces and Propagation Mechanisms. *manuscript, Arizona State University*.
49. McGrattan, E., Richard Rogerson, and Randall Wright (1997). An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy. *International Economic Review* 38, pp. 267-290.
50. Mendoza, E., Asaf Razin, and Linda Tesar (1994). Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption. *Journal of Monetary Economics* 34, pp. 297-323.
51. Naaman, M., Trang Dinh, and Jonathon Taylor (2008). Netflix Group Project. *manuscript, Rice University*.
52. Németh, B., Gábor Takács, István Pilászy, and Domonkos Tikk (2007). Major components of the gravity recommendation system. *ACM SIGKDD Explorations Newsletter* 9: 80.
53. Németh, B., Gábor Takács, István Pilászy, and Domonkos Tikk (2007). ["On the Gravity Recommendation System"](http://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings/gravity-Tikk.pdf) (PDF), *Proc. KDD Cup Workshop at SIGKDD*, San Jose, California, pp. 22–30, <http://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings/gravity-Tikk.pdf>.

54. Ohanian, L., Andrea Raffo, and Richard Rogerson (2008). Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries 1956-2004. *Journal of Monetary Economics*.
55. Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120–126.
56. Pencavel, John (1986), “Labor Supply of Men: A Survey,” in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, Vol. 1, North-Holland, Amsterdam, pp. 3-102.
57. Paterek, A. (2007). Improving Regularized Singular Value Decomposition for Collaborative Filtering. *KDD-Cup and Workshop*, ACM press.
58. Prescott, E. (2004). Why Do Americans Work So Much More than Europeans?. *Quarterly Review of the Federal Reserve Bank of Minneapolis*, pp. 2-13.
59. Prescott, E. (2006). Nobel Lecture: The Transformation of Macroeconomic Policy and Research. *Journal of Political Economy* 114, pp. 203-235.
60. Ragan, K. (2005). Taxes, Transfers and Time Use: Fiscal Policy in a Household Production Model. *manuscript, University of Chicago*.
61. Riedl, J. (2006). ClustKNN: a highly scalable hybrid model-& memory-based CF algorithm. In Proc. of WebKDD'06: KDD Workshop on Web Mining and Web Usage Analysis, at 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Philadelphia, PA, USA.
62. Rogerson, R. (2006). Understanding Differences in Hours Worked. *Review of Economic Dynamics* 9, pp. 365-409.

63. Rogerson, R. (2007). Taxes and Market Work: Is Scandinavia an Outlier?.
Economic Theory 32, pp. 59-85.
64. Rogerson, R. (2008). Structural Transformation and the Deterioration of European Labor Market Outcomes. *Journal of Political Economy* 116, pp. 235-259.
65. Rogerson, R., and Johanna Wallenius (2008). Micro and Macro Elasticities in a Life Cycle Model with Taxes. *NBER Working Papers 13017*, National Bureau of Economic Research, Inc.
66. Rogerson, R. (2009). Market Work, Home Work, and Taxes: A Cross-Country Analysis. *Review of International Economics*, Blackwell Publishing, vol. 17(3), pages 588-601.
67. Roweis, S. (1997). EM Algorithms for PCA and SPCA. *Advances in Neural Information Processing Systems* 10, pp. 626–632.
68. Rupert, P., Richard Rogerson, and Randall Wright (1995). Estimating Substitution Elasticities in Household Production Models. *Economic Theory* 6, pp. 179-193.
69. Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society B* 58.
70. Wang, J., A. P. de Vries and M. J. T. Reinders (2006). Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion. *Proc. 29th ACM SIGIR Conference on Information Retrieval*, pp. 501–508.